



UNIVERSITÀ DI NAPOLI
FEDERICO II

TESI DI DOTTORATO
IN SCIENZE SOCIALI E STATISTICHE
XXX CICLO

**Costrutti di PLS-PM di alto ordine
per l'analisi di indicatori sociali
compositi**

**Un'applicazione alla Customer satisfaction per la riduzione del
questionario**

Roberto Galasso

Una tesi presentata nel rispetto dei requisiti

per il

dipartimento di Scienze Sociali e Statistiche



Costrutti di PLS-PM di alto ordine per l'analisi di indicatori sociali compositi

Un'applicazione alla Customer satisfaction per la riduzione del questionario

Author
Roberto Galasso

Tutor
M. Gabriella Grassia

Co-tutor
Enrica Amatore

7 giugno 2018

INDICE

<i>Introduzione</i>	12
<i>1. La ricerca sociale e gli indicatori</i>	15
1.1 Introduzione	15
1.2 La selezione delle unità di analisi	17
1.3 Concetti e indicatori	19
1.4 Costuire gli indici	22
1.5 Validità e Attendibilità degli indici	23
1.6 Conclusioni	24
<i>2. Gli Indicatori Compositi</i>	25
2.1 Introduzione	25
2.2 Definizione di Indicatore Composito	26
2.3 Dagli Indicatori Compositi Data Driven agli Indicatori Compositi Model Based	28
2.3.1 Indicatori Compositi Theory Based e Data Driven	29
2.3.2 Indicatori Compositi Model Based	31
<i>3. Partial Least Squares Path Modeling</i>	34
3.1 Introduzione	34
3.2 Il PLS-PM	35
3.2.1 Modello di misura	35
3.2.2 Modello strutturale	40
3.3 L'algoritmo dei minimi quadrati parziali	40
3.3.1 Prima fase: Approssimazione iterativa delle variabili latenti	41
3.3.2 Seconda fase: Stima dei punteggi delle VL	45
3.3.3 Terza fase: Stima dei path coefficient	45
3.4 Validazione del modello	46
3.4.1 Valutazione dei risultati dei modelli di misura riflessivi	47
3.4.2 Valutazione dei risultati dei modelli di misura formativi	49

3.4.3	Valutazione dei risultati dei modelli strutturali	51
3.5	L'importanza dei punteggi degli Indicatori Compositi	55
3.6	Software disponibili per il PLS Path Modeling	56
4.	<i>Costrutti di alto ordine nel PLS-PM</i>	58
4.1	Introduzione	58
4.2	Stima dei costrutti nei modelli di ordine superiore	60
4.2.1	Modelli Molecular and Molar per costrutti di alto ordine	60
4.2.2	Tipi di modelli per costrutti di alto ordine	60
4.3	Approcci basati sul PLS-PM per la stima di modelli di alto ordine	62
4.3.1	L'approccio ad indicatori ripetuti	62
4.3.2	Approccio a due step	65
4.4	Nuovi metodi per costrutti di alto ordine nel PLS-PM	68
4.4.1	Approccio Misto	69
4.4.2	Approccio con regressione PLS	72
5.	<i>Studio di simulazione</i>	78
5.1	Generazione dati	78
5.2	Risultati: Modello a 3 variabili latenti di primo ordine con 8 manifeste per blocco	81
5.2.1	Confronto comunaltà	82
5.2.2	Path coefficients	83
5.2.3	Relative Bias	88
5.3	Risultati: Modello a 4 variabili latenti di primo ordine con 5 manifeste per blocco	90
5.3.1	Confronto comunaltà	90
5.3.2	Path Coefficient	90
5.3.3	Relative Bias	96
5.4	Risultati: Modello a 7 variabili latenti di primo ordine con 4 manifeste per blocco	99
5.4.1	Confronto comunaltà	99
5.4.2	Path Coefficient	101
5.4.3	Relative Bias	105
5.5	Conclusioni	109
6.	<i>Customer Satisfaction</i>	111
6.1	Introduzione	111
6.2	Perché rilevare la Customer Satisfaction	112

6.3	Come rilevare la Customer Satisfaction	113
6.4	Come costruire il questionario	115
6.4.1	Le scale di misurazione	117
6.5	Principali limiti del Questionario	120
6.6	NPS - Net Promoter Score	123
6.7	Conclusioni	124
7.	<i>Caso applicativo su un'azienda MultyUtility</i>	126
7.1	Applicazione dell'NPS	127
7.2	Applicazione della Customer Satisfaction	129
7.2.1	Il questionario	129
7.2.2	Il trattamento dei dati	131
7.2.3	L'analisi del modello di customer satisfaction con PLS- PM: Validazione	131
7.2.4	L'analisi del modello di customer satisfaction con PLS- PM: Interpretazione	135
7.2.5	Simulazione di applicazione	140
7.3	Conclusioni	141
	<i>Conclusioni</i>	142
	 <i>Appendice</i>	144
A.	<i>ALLEGATI</i>	145
A.1	Codice R per le simulazioni	146
A.2	Codice R per le simulazioni sull'Applicazione	161
	<i>Bibliografia</i>	167

ELENCO DELLE FIGURE

1.1	Schema di ricerca	16
1.2	Processo di creazione di un indice	18
3.1	Path-diagram di un modello riflessivo	35
3.2	Path-Diagram di un modello formativo	38
3.3	Path-Diagram del modello strutturale	40
3.4	Analisi della ridondanza per la validità convergente	50
3.5	Grafico Swot	56
4.1	Tipi di costrutti di alto ordine	61
4.2	Approccio ad indicatori ripetuti	63
4.3	Approccio a due step	65
4.4	Costrutto di secondo ordine con tutte le MV del primo ordine	70
4.5	Punteggi di ogni singolo costrutto di basso ordine	71
4.6	Costrutto di secondo ordine con i punteggi del primo ordine	71
4.7	Costrutto di secondo ordine con tutte le MV del primo ordine	76
4.8	Costrutto di secondo ordine con le componenti PLS-R del primo ordine	77
5.1	Path Diagram	79
5.2	Confronto Comunalità	83
5.3	Confronto Comunalità	90
5.4	Confronto Comunalità	99
7.1	Percentuali totali	127
7.2	Percentuali per tipo cliente	128
7.3	Percentuali per area geografica	128
7.4	Risultati Modello Pls Classico	135
7.5	Risultati Modello Pls Gerarchico	136
7.6	Punteggi medi delle variabili latenti e relativi impatti sulla <i>Soddisfazione</i> con modello classico.	137
7.7	Punteggi medi delle variabili latenti e relativi impatti sulla <i>Soddisfazione</i> con modello gerarchico.	137

7.8	Medie delle variabili manifeste e relativi loadings	139
-----	---	-----

ELENCO DELLE TABELLE

5.1	Confronto Comunalità	82
5.2	Modello Formativo-Formativo	84
5.3	Modello Riflessivo-Riflessivo	85
5.4	Modello Formativo-Riflessivo	86
5.5	Modello Riflessivo-Formativo	87
5.6	Relative Bias Modello Formativo-Formativo	88
5.7	Relative Bias Modello Riflessivo-Riflessivo	88
5.8	Relative Bias Modello Riflessivo-Formativo	89
5.9	Relative Bias Modello Formativo-Riflessivo	89
5.10	Confronto Comunalità	91
5.11	Modello Formativo-Formativo	92
5.12	Modello Riflessivo-Riflessivo	93
5.13	Modello Formativo-Riflessivo	94
5.14	Modello Riflessivo-Formativo	95
5.15	Relative Bias Modello Formativo-Formativo	96
5.16	Relative Bias Modello Riflessivo-Riflessivo	97
5.17	Relative Bias Modello Riflessivo-Formativo	97
5.18	Relative Bias Modello Formativo-Riflessivo	98
5.19	Confronto Comunalità	100
5.20	Modello Formativo-Formativo	101
5.21	Modello Riflessivo-Riflessivo	102
5.22	Modello Formativo-Riflessivo	103
5.23	Modello Riflessivo-Formativo	104
5.24	Relative Bias Modello Formativo-Formativo	105
5.25	Relative Bias Modello Riflessivo-Riflessivo	106
5.26	Relative Bias Modello Riflessivo-Formativo	107
5.27	Relative Bias Modello Formativo-Riflessivo	108
6.1	Tasso di abbandono per numero di pagine	123
6.2	Tasso di abbandono per minuti di compilazione	123
7.1	Questionario di Soddisfazione del cliente	130

7.2	Verifica dell'unidimensionalità dei blocchi Modello Classico.	132
7.3	Verifica dell'unidimensionalità dei blocchi Modello Gerarchico.	132
7.4	Tabella dei pesi e delle correlazioni per il Modello Classico .	133
7.5	Tabella dei pesi e delle correlazioni per il Modello Gerarchico	134
7.6	Punteggi medi delle variabili latenti	136
7.7	Percentuali di ben specificati con 5 VL e 5 VM per blocco . .	141

INTRODUZIONE

Il presente progetto di ricerca si focalizza sullo studio, il confronto e l'utilizzo di indicatori compositi attraverso i modelli ad equazioni strutturali, nello specifico attraverso l'uso dei modelli *Partial Least Squares - Path Modeling*. Molti fenomeni sociali, economici e psicologici sono complessi e spesso difficili da misurare e valutare. La complessità implica la multidimensionalità e un alto livello di astrazione. I ricercatori negli ultimi anni si sono focalizzati sullo sviluppo e l'uso di particolari indicatori proprio con l'obiettivo di ottenere una descrizione globale di fenomeni complessi e per avere una sintesi adeguata di informazioni. Gli indicatori compositi hanno ricevuto notevole attenzione negli ultimi anni e sono state sviluppate varie metodologie per gestirne diversi aspetti. Saisana definisce un indicatore come "una combinazione matematica di singoli indicatori che rappresentano diverse dimensioni di un concetto la cui descrizione è l'obiettivo dell'analisi". Pertanto la caratteristica principale di un indicatore composito è quella di riassumere questioni complesse e multidimensionali. La costruzione di un indicatore composito implica la ricerca di una sintesi adeguata di un certo numero di variabili osservate, dette anche variabili manifeste, al fine di ottenere una semplice rappresentazione di un fenomeno multidimensionale. Un indicatore composito può essere considerato come un concetto latente, non direttamente misurabile, la cui stima può essere ottenuta attraverso indicatori elementari. Gli indicatori compositi sono molto utili per affrontare quei fenomeni complessi che non possono essere osservati direttamente. Diversi concetti sociali ed economici, come ad esempio la qualità della vita, la soddisfazione o la coesione sociale, rappresentano concetti multidimensionali che sono difficili e complessi da definire. Il problema principale per i ricercatori è quello di trovare strumenti adeguati per ottenere un indicatore composito. La letteratura esistente offre diversi metodi alternativi per ottenere un indicatore composito. Questa ricerca si focalizza sul testare, confrontare ed infine utilizzare per nuove applicazioni, dei particolari metodi quali i Modelli ad Equazioni Strutturali (*Structural Equation Model* - SEM) e, nello specifico, un approccio *Partial Least Squares - Path Modeling* (PLS-PM). Grazie a specifiche caratteristiche, i SEM rappresenta-

no uno strumento utile per costruire sistemi di indicatori compositi. Tali modelli stimano le relazioni casuali tra diversi costrutti latenti ciascuno dei quali è collegato ad una serie di indicatori osservabili, di variabili manifeste, attraverso una relazione formativa, riflessiva o mista. I parametri di ogni SEM possono essere stimati seguendo due diversi approcci:

- l'approccio basato sulla covarianza tra le variabili manifeste, i cosiddetti *covariance-based methods*
- l'approccio basato sulla ricerca di particolari componenti latenti, i cosiddetti *component-based methods*.

Il PLS-PM è un *component based method*, cioè è un approccio basato sulla ricerca di componenti latenti. Lo scopo è quello di identificare delle variabili latenti che siano allo stesso tempo il più rappresentative possibile delle loro variabili manifeste e che siano in grado di spiegare al meglio le relazioni tra blocchi così come definite dalla struttura del modello interno. Ogni indicatore composito è ottenuto in modo da essere il più rappresentativo dei suoi indicatori sottostanti e il più correlato con gli altri indicatori compositi a cui esso è collegato. Il PLS-PM è un algoritmo iterativo che permette di stimare le relazioni esistenti tra i diversi blocchi di variabili attraverso un sistema di equazioni lineari. Negli ultimi anni la ricerca si è focalizzata su un particolare aspetto legato alla multidimensionalità e ad un alto livello di astrazione. Infatti, l'approccio PLS-PM può essere utilizzato per studiare modelli con un alto livello di astrazione, anche quando la variabile latente è complessa e manca di proprie variabili manifeste e viene descritta da diversi blocchi sottostanti. Il disegno originale del PLS-PM di Wold prevedeva che ogni costrutto fosse necessariamente collegato ad un set di variabili osservate. Su questa base, Lohmöller nel 1989 propone una procedura per trattare costrutti gerarchici, il cosiddetto modello a componente gerarchica. I costrutti gerarchici, detti anche multidimensionali, sono costrutti che coinvolgono più di una dimensione e possiamo distinguerli dai costrutti unidimensionali, i quali sono caratterizzati da una singola dimensione sottostante. I costrutti di ordine superiore sono rappresentazioni esplicite di costrutti multidimensionali che esistono ad un livello superiore di astrazione e sono legati ad altri costrutti sottostanti.

In questo lavoro, dopo aver presentato vari approcci al PLS-PM di alto ordine, sia già noti in letteratura sia nuovi sviluppi a cui ho collaborato durante gli anni di dottorato di ricerca, è stato approfondito lo sviluppo di questi modelli per l'adattamento a tutti i tipi di relazioni tra variabili latenti e variabili manifeste, sia nel modello strutturale che nel modello di misura. Que-

sta tematica in letteratura risulta ancora molto difficile da valutare e viene affrontata in questo lavoro con delle prime simulazioni che mettono a confronto i metodi già presenti in letteratura con i metodi di recente sviluppo. Partendo da questi studi, si è giunti ad un ulteriore utilizzo di tali modelli. Sappiamo che in alcuni ambiti socio-economici, spesso, i dati sono raccolti tramite questionario. La somministrazione dei questionari risulta a volte complessa per questioni di tempo e di denaro. Diventa dunque importante ridurre la lunghezza di questo strumento di rilevazione, così da ridurre tempi, costi e tassi di abbandono. Si presenta dunque un'applicazione del PLS-PM di alto ordine sulla Customer Satisfaction per la riduzione della lunghezza del questionario. Utilizzando un questionario classico con domande relative anche alla soddisfazione generale del cliente, si è applicato un modello PLS-PM classico (utilizzando le variabili manifeste della variabile di alto ordine *Soddisfazione*) e un modello PLS-PM gerarchico (non considerando le variabili manifeste del blocco di alto ordine). L'obiettivo è stato quello di dimostrare che, eliminando le variabili manifeste relative alla soddisfazione e, di conseguenza, riducendo la lunghezza del questionario, può essere applicato un modello gerarchico, il quale produce risultati simili a quelli che ottenuti utilizzando un modello classico, sia per quanto riguarda la validazione del modello che per interpretazione degli stessi.

1. LA RICERCA SOCIALE E GLI INDICATORI

1.1 Introduzione

La ricerca empirica è un processo conoscitivo che partendo da alcuni interrogativi, arriva a fornire ad essi risposte adeguate [3]. Il ricercatore deve trovare, tra i tanti possibili, il modo più adatto a fornire le risposte, utilizzando metodi trasparenti e riproducibili che rendano verificabili tali risposte. Per poter affrontare un progetto di ricerca empirica in maniera adeguata è necessario innanzitutto progettare le fasi del lavoro. Questa fase preliminare è definita *disgeno di ricerca*. Essa è indispensabile qualunque sia il tipo di approccio che si sceglie di seguire e durante questa fase vanno definiti:

- Gli obbiettivi di ricerca;
- Il contesto della ricerca;
- I concetti di indagine;
- I metodi di rilevazione;
- Il tipo di studio/analisi che si intendono effettuare;
- Le unità minime di analisi ;
- Il ruolo del ricercatore.

Nel pianificare un *disgeno di ricerca* è ovvio che il ricercatore si basi sul bagaglio teorico proprio e della propria disciplina. Riprendendo dunque Amaturò : [3] “Si può affermare che se quando ci si muove sul piano teorico ci si esprime utilizzando concetti tra cui si individuano relazioni causali, espresse in asserti di carattere generale, quando si passa al piano dell’empiria bisognerà esprimersi utilizzando indicatori, tra cui si ipotizzano correlazioni che vengono controllate su casi particolari. Per governare questo passaggio vanno quindi affrontate tre questioni principali: quella della operazionalizzazione o definizione operativa dei concetti (tema definito per

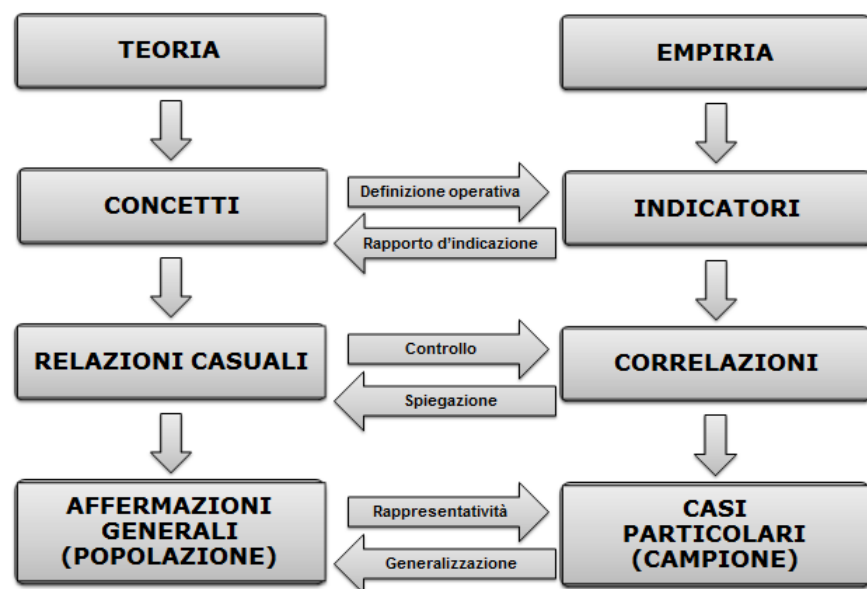


Fig. 1.1: Schema di ricerca

molto tempo della misura), quella del controllo e quella della rappresentatività. [...] Dunque qualunque tipo di conoscenza si fonda su costrutti concettuali tra i quali si ipotizzano relazioni”.

In un ambito scientifico è necessario che ogni idea venga riportata ad un valore empirico. Il concetto, perciò, andrà suddiviso e ridotto in una serie di concetti più semplici e, soprattutto, di cui sia facile determinare un valore empirico.

Sarà poi necessario analizzare empiricamente se e quanto i concetti più semplici siano associati tra di loro e con il concetto di analisi.

È in questi punti cruciali della ricerca che la statistica e le varie discipline di ricerca si intrecciano, dando al ricercatore la possibilità di utilizzare metodi che individuino i rapporti di causalità tra i concetti complessi e quelli semplici.

Infine, sarà necessario scegliere in maniera adeguata le singole unità da analizzare in modo che esse rappresentino al meglio tutto l’ambito di riferimento.

Come ovvio aspettarsi, ogni materia o ambito di ricerca ha le proprie peculiarità e ciò induce il ricercatore a utilizzare stili di ricerca differenti. Essi possono essere riportati a due grandi categoria di ricerca: la ricerca *quali-*

tativa e la ricerca *quantitativa*. La prima utilizzata maggiormente in campo psicologico, epidemiologico e di recente in ambito sociologico è definita da Bryman e Burgess [15] come : “un processo dinamico che lega assieme problemi, teorie e metodi [...] di conseguenza il processo di ricerca non è una ben definita sequenza di procedure che seguono il nitido disegno, ma una confusa interazione tra il mondo concettuale e quello empirico, dove deduzione ed intuizione si realizzano nello stesso tempo”.

La ricerca quantitativa (o standard) è invece caratterizzata dalla bassa interazione con gli elementi di analisi, così da minimizzare il rischio di “inquinare” i dati da parte del ricercatore. La raccolta, il trattamento dei dati, l’impiego dei dati e l’uso della statistica seguono dei protocolli definiti e replicabili. Questo modo di procedere parte dagli studi di Lazarsfeld [12];[100];[101] a cui viene dato il nome di linguaggio delle variabili [3]. Si definisce dunque uno stile di ricerca che si fonda sullo studio delle relazioni tra concetti tradotti in variabili numeriche di cui è possibile determinare un valore empirico.

Volendo ulteriormente schematizzare i passaggi di una ricerca con approccio quantitativo, possiamo partire dal concetto che si ritiene l’oggetto dell’analisi e delle proprietà che si possono riferire ad esso. Definito l’oggetto e le sue proprietà, sarà necessario capire qual è l’unità di osservazione e identificare un indicatore che sia un valore empirico delle sue proprietà. Ogni oggetto identificato da un’unità di osservazione verrà definito come caso (o osservazioni) e i suoi indicatori vengono definiti variabili. Una volta effettuata la rilevazione di tali variabili si otterrà una matrice dei dati che conterrà n righe ed m colonne, che saranno rispettivamente gli n casi e le m variabili. Dall’analisi (statistica o meno) di tale matrice dei dati verrà calcolato un indice che può essere definito come un valore riassuntivo del concetto di analisi.

1.2 La selezione delle unità di analisi

Nell’individuare le unità di analisi in primis è necessario definire l’oggetto al quale le proprietà del concetto si riferiscono. La ricerca quantitativa studiando le relazioni tra le proprietà dell’oggetto e l’oggetto stesso non può prescindere da una accurata selezione dell’unità minima su cui basare la ricerca stessa. L’oggetto della ricerca sociale non è necessariamente un individuo. I vari oggetti di analisi sono chiamati unità di analisi o di osservazione. Si definisce l’unità di analisi o di osservazione “il tipo di oggetto di ricerca di cui si occupa una ricerca; è il referente sul quale il

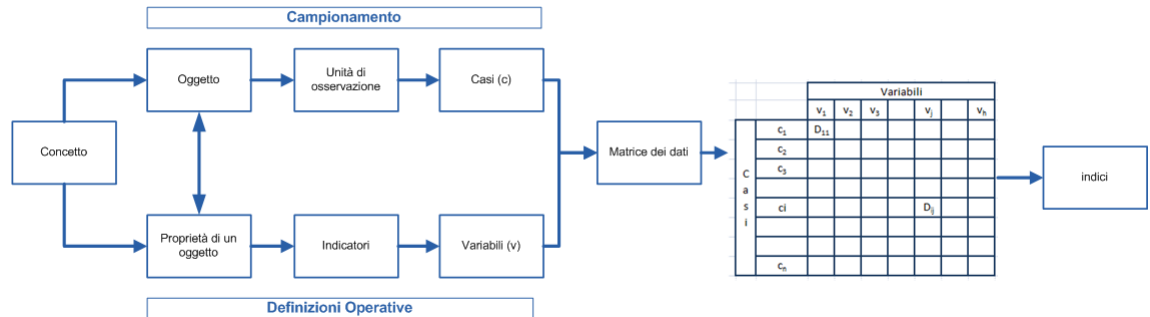


Fig. 1.2: Processo di creazione di un indice

ricercatore intende rilevare alcune informazioni; essa è singolare ed astratta” [3]. Una volta selezionata l’unità di analisi essa si identifica come *caso*, quindi partendo dal più alto livello di astrazione è possibile distinguere tra oggetto, unità e caso di ricerca. Nell’individuare i casi concreti su cui effettuare la ricerca è fondamentale definire un ambito territoriale ed un periodo temporale a cui riferirsi, in quanto ogni oggetto di ricerca è mutevole nello spazio e nel tempo. La fase di individuazione dei casi deve dunque seguire delle regole e dei criteri precisi che permettano di ottenere dei casi di analisi che ben rappresentino l’oggetto di analisi. Questo insieme di criteri viene definito campionamento. I risultati ottenuti dalla ricerca divengono quindi generalizzabili solo all’interno dell’ambito definito in fase di campionamento.

Esistono vari tipi di unità di analisi a cui è possibile riferirsi nell’ambito della ricerca. In generale si possono distinguere: le unità di analisi o di riferimento e le unità di raccolta o di rilevamento. Le unità di raccolta si definiscono come le unità sulle quali vengono raccolte informazioni, mentre le unità di analisi sono le unità di cui vogliamo conoscere l’informazione. A questi due tipi di unità bisogna aggiungere l’unità statistica che rappresenta l’unità elementare su cui si condurranno le analisi statistiche.

Su questa base è possibile distinguere quattro tipi di unità:

- L’individuo è l’unità a cui la maggior parte delle indagini sociali fa riferimento, non è scomponibile e costituisce l’unità elementare delle unità aggregate.
- L’aggregato o collettivo, si può scomporre in unità più piccole. Nella

ricerca sociale si identificano tre tipi di aggregati. Il primo è caratterizzato dalla presenza dell'elemento individuale, in questo caso le unità di analisi e di raccolta non coincidono, perché le informazioni vengono rilevate individualmente all'interno del gruppo e poi estese al collettivo. Nel secondo tipo di unità aggregate le informazioni vengono rilevate a livello territoriale, si tratta infatti di unità ecologiche o territoriali. Infine l'ultimo tipo di unità aggregata è l'ente, cioè organizzazioni e istituzioni con una propria struttura e regolamentazione, in questo caso l'unità di rilevamento è rappresentata dal collettivo stesso.

- Evento sociale è un'unità che fa riferimento ad un avvenimento occasionale.
- Prodotto culturale o rappresentazione simbolica è l'unità presa in considerazione se l'analisi ha il fine di studiare un fenomeno sociale.

1.3 Concetti e indicatori

Entrando nel dettaglio degli indicatori è necessario chiarire alcuni punti. Come detto in precedenza, per procedere con un'analisi qualitativa è necessario tradurre un concetto in una variabile attraverso l'attribuzione di un valore misurabile empiricamente.

A questo processo di traduzione si dà il nome di operativizzazione [3]; si dice anche che di un determinato concetto viene data una definizione operativa.

All'interno di questo processo bisogna stabilire con che tipo di valore vogliamo misurare la variabile (valori continui, valori discreti, categorizzazioni ecc., si veda anche il paragrafo 6.4.1)

È inoltre necessario decidere come ricavare questo valore, ad esempio, tramite il contatto diretto con l'unità di analisi (si approfondirà largamente l'argomento nei capitoli 6 e 7) o desumendola da altre fonti quali istituti nazionali o altre indagini ecc.

Ognuna di queste opzioni può essere più o meno adatta a seconda del tipo di ricerca che verrà effettuata, dalla popolazione di riferimento o dall'obiettivo preposto e può essere influenzata e indotta da limiti fisici, tecnologici o di budget.

In molti campi della ricerca il principale limite sta appunto nella difficoltà di trovare un valore empirico in modo diretto o semplice, per dei

concetti che hanno un significato ampio e complesso o, in alcuni casi, completamente astratto.

Quindi, quando i concetti sono così generici che non è possibile definirne delle proprietà direttamente osservabili, è necessario semplificare tale concetto con uno o più concetti il cui valore sia misurabile empiricamente. Questo concetto misurabile viene definito come indicatore (I) ed esso è una proprietà che fornisce informazioni su un altro concetto (C) più generale.

Un indicatore è dunque un concetto più specifico, che definisce in maniera più o meno approssimata il concetto più generale oggetto della ricerca.

Un indicatore in generale è sempre una condizione manifesta e dunque misurabile, che da informazioni su qualcosa di non manifesto; inoltre esso può essere di interesse per la ricerca sia come approssimazione del concetto generale sia in quanto condizione a se stante ed, infine, un indicatore può essere considerato come approssimazione di vari fenomeni a seconda del punto di vista e dello scopo finale del ricercatore.

Scegliere un indicatore significa, come già detto, approssimare il concetto latente, questo crea un “apporto di indicazione” (o rappresentazione semantica) [113];[115] tra il concetto e l’indicatore. Marradi [115] identifica inoltre la parte indicante come la parte di contenuto semantico dell’indicatore, presente anche nel concetto. Vedremo nei capitoli successivi come, attraverso tecniche statistiche, sia possibile stimare la parte indicante così da comprendere se gli indicatori scelti siano validi al fine di rappresentarne il concetto. Quindi più la parte indicante è ampia tanto più l’indicatore è un valido rappresentante del concetto.

Lazarsfeld [100] descrisse come costruire variabili capaci di misurare oggetti complessi seguendo un processo tipico che permette di esprimere i concetti in termini di indici e che si compone di quattro fasi principali. La prima fase è quella della rappresentazione figurata del concetto. La riflessione e l’analisi, che precedono la costruzione di uno strumento di misurazione, nascono da una rappresentazione figurata. Il ricercatore analizzando i dettagli di un problema teorico abbozza per prima cosa una costruzione astratta. L’aspetto creativo del suo lavoro comincia nel momento in cui, dopo aver percepito fenomeni diversi, cerca di scoprire un aspetto caratteristico fondamentale e spiega le eventuali regolarità che osserva. Il concetto è soltanto un’entità concepita in termini vaghi che dà senso alle relazioni osservate tra i fenomeni.

La fase successiva, definita specificazione del concetto, consiste nell’analisi delle “componenti” dei concetti originari che vengono definiti, secondo i casi, “aspetti” o “dimensioni”. Esse possono essere dedotte anali-

ticamente dal concetto generale o empiricamente dalla struttura della loro interazione. In ogni caso, un concetto corrisponde generalmente a un insieme complesso di fenomeni piuttosto che a un fenomeno semplice e direttamente osservabile. La teoria della misura [100] attribuisce a questi fattori il nome di “dimensioni”, la cui analisi è spesso un problema complesso. Definire tutte le dimensioni di un concetto potrebbe non essere possibile ma, in generale, la complessità dei concetti impiegati in sociologia è tale che la loro specificazione preveda più dimensioni.

La terza fase consiste nella scelta degli indicatori per le dimensioni considerate. Non mancano anche qui le difficoltà. La prima si può formulare così: che cos’è esattamente un indicatore? Dalle parole di William James [80] potremmo sintetizzare che un indicatore è “un modo pratico di esprimere astrattamente un aspetto comune” di un concetto. Oggi si tende a specificare la relazione di questi indicatori con l’idea fondamentale che non si può coprire tutto il concetto di ricerca. Si parla dunque di probabilità che un concetto sia identificato da vari indicatori. Dati questi presupposti durante la selezione degli indicatori non bisogna dimenticare che gli indicatori a disposizione del ricercatore variano ampiamente a seconda dei contesti di analisi, ed è quindi importante elaborare un indice che tenga conto anche del contesto. Inoltre proprio poiché la relazione tra ogni indicatore e il suo concetto è definita in termini di probabilità, è indispensabile utilizzare un numero elevato di indicatori.¹

La quarta fase consiste nella Formazione degli indici. Dopo aver scomposto in dimensioni e scelto gli indicatori per ogni dimensione, bisogna costruire un unico indice partendo da queste informazioni elementari. Per poter confrontare tra loro le unità di analisi o interpretare facilmente un concetto, è necessario generare un indice generale che copra l’insieme dei dati. Anche nei casi in cui si vuole indagare la relazione tra le dimensioni di analisi e variabili esterne, una sintesi dei diversi indicatori in un unico indice è utile per l’interpretazione dei concetti di ricerca. Come già accennato, formalmente, ogni indicatore mantiene un rapporto di probabilità con la variabile oggetto di studio. Ciò fa sì che un’unità di analisi non muti la sua posizione rispetto ad un concetto, anche se un indicatore si modifica sensibilmente. È anche possibile che tale posizione cambi senza che un indicatore muti in maniera particolare, ma che ci siano stati minimi spostamenti in tutti gli indicatori. Se un indice raccoglie in sé molti indicatori ciò diviene meno probabile. Esiste un’ampia fetta della ricerca, soprattutto

¹ Nella letteratura statistica si vedrà che questo è vero fino ad un limite di indicatori oltre il quale non si aggiungono informazioni rilevanti

to statistica, che si occupa di trovare risposte a queste problematiche. Nei capitoli successivi se ne affronteranno alcune.

1.4 *Costuire gli indici*

Come visto, alla base della teoria di Lazarsfeld [10] abbiamo l'idea che gli indicatori siano una rappresentazione parziale del concetto, tali indicatori vengono quantificati attraverso le variabili. L'ultimo step del processo è dunque la costruzione degli indici che consentano la sintetizzazione di più variabili e che vadano essi stessi a rappresentare una nuova variabile che riassume tutte le informazioni che sono state prese in considerazione durante l'analisi. In sintesi un indice aiuta a ricomporre l'unità del concetto.

La costruzione degli indici, dunque, costituisce l'ultima fase del processo di passaggio da un concetto alla variabile corrispondente: si tratta di sintetizzare la pluralità delle variabili in cui vengono tradotti gli indicatori che rappresentano semanticamente un concetto complesso. Da essa si ottiene appunto l'indice. Quest'ultimo non è altro che una nuova variabile che raggruppa in sé tutte le informazioni raccolte da ciascuna variabile sulle unità di analisi. Quando infatti un concetto viene scomposto nelle sue dimensioni, come già accennato in precedenza, diventa forte l'esigenza di ottenere una sintesi, che ci permetta di ricostruire in qualche modo l'unità del concetto di interesse. Questa necessità è particolarmente forte quando lo scopo del ricercatore è di utilizzare il concetto in questione come variabile, dipendente o indipendente, in una relazione di causa-effetto: diventerebbe infatti troppo complicato dal punto di vista statistico analizzare un modello con troppe variabili. Nei capitoli successivi si discuterà proprio sulla possibilità di utilizzare gli indici come indicatori all'interno di altri modelli.

Esistono due tipi di indici: tipologici e sommatori. Quando partiamo da variabili categoriali (ordinate e non) è possibile costruire un indice tipologico attraverso l'utilizzo della tabella di contingenza, nella quale le categorie di una variabile sono riportate nelle righe e le categorie dell'altra sono riportate nelle colonne, e in ogni cella c'è un tipo, cioè la combinazione di una categoria della variabile nelle righe con la categoria della variabile nelle colonne. Per scegliere i diversi tipi è necessario seguire dei criteri logici che devono rispettare le seguenti regole:

- Ciascun tipo deve essere esaustivo e mutuamente esclusivo;
- Ciascun tipo deve avere omogeneità semantica interna;

- Nella costruzione dei tipi deve essere rispettato un certo equilibrio numerico.

Gli indici sommatori sono utilizzabili con variabili cardinali o quasi-cardinali e vengono costruiti quindi attraverso operazioni algebriche e matematiche anche notevolmente complesse.

1.5 Validità e Attendibilità degli indici

Come abbiamo affermato in precedenza, i concetti che si vogliono misurare sono spesso molto complessi, di conseguenza risulta complesso scegliere gli indicatori opportuni. Nasce quindi il problema di stabilire quali siano le caratteristiche che consentono di definire un “buon” indicatore, data la complessità dei fenomeni che sono da essi analizzati tali caratteristiche sono difficilmente dimostrabili. La scelta degli indicatori può comportare, ovviamente, degli errori che minerebbero l’esattezza dell’intera analisi. Un indicatore per essere capace di misurare il concetto oggetto di studio deve avere le seguenti caratteristiche:

- Validità, ovvero il grado con cui uno strumento misura ciò che afferma di misurare
- Attendibilità, cioè il grado di accordo o coerenza tra misurazioni indipendenti dello stesso costrutto. Se sono stati fatti molti errori, questa correlazione è bassa; se pochi, alta.

Per costruire un buon indicatore sono necessarie anche altre caratteristiche, come ad esempio Sensibilità (per discriminare le diverse forme che può assumere un fenomeno), Adeguatezza (per rispondere ai bisogni conoscitivi del ricercatore), Comparabilità (rispetto alla mutevolezza spazio temporale), ma le uniche due proprietà su cui è possibile effettuare delle verifiche con strumenti statistici sono la validità e l’attendibilità, tutto il resto è affidato al buon senso del ricercatore nonché alla difficoltà del concetto che si vuole misurare.

Anche in questo caso nei capitoli successivi saranno affrontati temi riguardanti la validità e l’affidabilità dei modelli. È infatti fondamentale nell’ambito della ricerca statistica definire dei valori confrontabili di validità e attendibilità, così da poter confrontare tra loro vari modelli di costruzione di uno stesso indice. In alcuni casi la letteratura statistica fornisce

metodi molto precisi e standard, ma nel caso dell'analisi della costruzione di indici (o indicatori) complessi non sempre questo tipo di confronti è "standardizzato" e in molti casi è ancora oggetto di studio.

1.6 Conclusioni

Essendo la ricerca empirica un processo conoscitivo che, partendo da alcuni interrogativi, arriva a fornire ad essi risposte adeguate, sta al ricercatore trovare, il modo più adatto a fornire delle risposte, utilizzando metodi trasparenti e riproducibili che rendano verificabili tali risposte. Nell'affrontare un progetto di ricerca empirica in maniera adeguata il ricercatore deve superare varie fasi del lavoro che vengono, in un primo momento, schematizzate nel disegno di ricerca, così da poter definire il più dettagliatamente possibile, gli obiettivi di ricerca, il contesto, i concetti di indagine, i metodi di rilevazione, il tipo di studio o analisi che si intendono effettuare e le unità minime di analisi. Durante queste fasi è necessario riportare ogni concetto ad un valore empirico, ma spesso tale concetto non è semplice e quindi non è misurabile direttamente. In questo contesto nasce quindi l'esigenza di scomporre concetti complessi in concetti semplici che siano misurabili. Sarà poi necessario analizzare empiricamente se e quanto i concetti più semplici siano associati tra di loro e con il concetto di analisi. È in questi punti cruciali della ricerca che la statistica e le varie discipline di ricerca si intrecciano, dando al ricercatore la possibilità di mettere in pratica metodi che individuino i rapporti di causalità tra i concetti complessi e quelli semplici. Su queste basi, si indagherà più a fondo in questa tesi, entrando nello specifico degli indicatori compositi e di come essi, essendo alla base di molte teorie sociologiche (e non solo), siano diventati forte oggetto di studio anche in statistica. In letteratura statistica è diffuso lo studio di tali indicatori, ma soprattutto di come sia possibile fornire ai ricercatori dei modelli validi e affidabili per sintetizzare concetti molto complessi, se non del tutto astratti. In conclusione la ricerca sociologica empirica (sia qualitativa che quantitativa) risulta fortemente connessa alla statistica che ne diventa ormai parte integrante, fornendo svariati strumenti di supporto.

2. GLI INDICATORI COMPOSITI

2.1 Introduzione

Il mondo reale è caratterizzato da una profonda complessità. Molti fenomeni socioeconomici sono complessi e, a volte, difficili da misurare e valutare. Un fenomeno è definito complesso quando gli aspetti rilevanti di un particolare problema non possono essere catturati usando solo una singola prospettiva [55]. È necessario considerare il concetto formato da differenti dimensioni, ognuna delle quali rappresenta un differente aspetto di esso e che interagisce con tutte le altre dimensioni. Per questa ragione, il più delle volte, la complessità implica anche la multidimensionalità [41], influenzando così la misurazione del fenomeno che stiamo analizzando. I risultati sono determinati non da singole causa, ma da moltedici esse che, la maggior parte delle volte, non interagiscono in modo additivo. Questo vuol dire che l'effetto combinato non è necessariamente la somma degli effetti separati. Negli ultimi anni, il dibattito sulla misurazione di fenomeni multidimensionali ha causato, all'interno della comunità scientifica mondiale, un rinnovato interesse grazie alla pubblicazione della relazione di Stiglitz del 2009 e la prima relazione sul "Benessere Equo e Sostenibile" (BES) del Comitato composto da ISTAT (Istituto Nazionale Statale Italiano) e CNEL (Consiglio italiano per l'economia e il lavoro) pubblicata nel 2013. È divenuta consapevolezza comune che alcuni fenomeni socio-economici devono essere considerati e rappresentati da più di una dimensione. Fenomeni come la povertà, lo sviluppo, il benessere, la qualità della vita, richiedono, per essere misurati, la combinazione di differenti dimensioni, una combinazione che può essere ottenuta applicando metodologie conosciute come *Indicatori Compositi* [118]; [159].

Una volta riconosciuta la multidimensionalità, la misurazione del fenomeno ha numerosi problemi teorici e metodologici che non sono presenti nel tradizionale approccio unidimensionale. Come già accennato nel primo capitolo, il primo problema riguarda la scelta delle dimensioni: quali e quante dimensioni sono rilevanti e devono essere considerate? Questo è anche chiamato da Sen il problema dell'appropriata "base informativa"

[171]. Inoltre, dobbiamo capire se esistono rapporti tra le varie dimensioni e, in caso affermativo, comprendere la loro natura. Pertanto, in una prospettiva multidimensionale e tenendo conto delle eventuali relazioni tra le dimensioni, si parla di *un sistema di indicatori compositi*, che misura e rappresenta differenti dimensioni del fenomeno osservato. Tale sistema di indicatori compositi non rappresenta una pura e semplice raccolta di indicatori, ma fornisce ai ricercatori informazioni che vanno oltre la semplice somma degli elementi.

2.2 Definizione di Indicatore Composito

Secondo Saisana et al. [127], “un *Indicatore Composito* (IC) è definito come una combinazione matematica di singoli indicatori che rappresentano diverse dimensioni di un concetto, la cui descrizione, è l’obiettivo dell’analisi. Esso si forma quando i singoli indicatori vengono compilati in un unico indice sulla base di un modello sottostante. Gli IC dovrebbero idealmente misurare un concetto multidimensionale che non può essere catturato da un singolo indicatore, come ad esempio la competitività, l’industrializzazione, la sostenibilità, l’integrazione del mercato unico, ecc.” [127]. La caratteristica principale di un IC è che riassume problemi complessi e multidimensionali. La costruzione di questo tipo di indicatore implica la ricerca di una sintesi adeguata di Variabili Manifeste (VM) per ottenere una semplice rappresentazione di un fenomeno multidimensionale. Di conseguenza, un IC, può essere considerato come un concetto latente, non direttamente misurabile, la cui stima può essere ottenuta attraverso il valore degli Indicatori Elementari (IE) o VM.

La costruzione di un IC e il suo utilizzo porta a definire una serie di vantaggi e svantaggi. Il principale vantaggio è che un IC può essere utilizzato per riassumere concetti multidimensionali e potrebbe aiutare a sintetizzare un elenco di indicatori. Invece, il problema più serio, è che essi possono mandare messaggi fuorvianti se sono mal costruiti o erroneamente interpretati. Tutti i pro e i contro sono trattati in Saisana et al. [127]. Per superare questi problemi, la letteratura si focalizza, per la costruzione di un IC, su una serie di passaggi fondamentali:

- Decidere il fenomeno da studiare e capire se necessita dell’uso di un IC;

- Selezionare degli IE. È necessario capire quali sono gli indicatori più rilevanti per misurare il fenomeno. Non esiste un modo pienamente oggettivo per la selezione degli indicatori rilevanti;
- Valutare la qualità dei dati. È necessario che i dati di tutti gli indicatori siano di alta qualità;
- Valutare le relazioni tra gli indicatori. Metodi come l'Analisi delle Componenti Principali possono fornire indicazioni sulle relazioni tra gli IE. Questo è considerato un prerequisito per l'analisi preliminare degli IE;
- Normalizzare e ponderare gli indicatori. Esistono molti metodi in letteratura per la normalizzazione e ponderazione degli IE. La scelta del metodo più appropriato dipende dai dati e dall'analista;
- Testare la robustezza e la sensibilità. Variazioni inevitabili nel sistema di ponderazione e nella scelta degli IE influenzeranno i risultati dell'IC.

Ogni passaggio è estremamente importante, ma la coerenza in tutto il processo è altrettanto vitale. Le scelte realizzate in un solo passaggio possono avere importanti implicazioni per gli altri passaggi. In letteratura sono presenti molti lavori di approfondimento sulla costruzione degli IC, da segnalare un lavoro dell' OCSE [127] che offre delle *"recommended practices"* per la costruzione degli IC [157] inoltre in Nando et al. [127] si discute sui diversi passaggi per la costruzione degli IC, insieme ai pro e contro legati all'utilizzo di informazioni statistiche aggregate.

Va però almeno fatto un piccolo approfondimento sulle relazioni degli IC e degli IC con i loro IE. Questi tipi di relazioni modificano infatti il concetto stesso della formazione dell'indice e fanno riferimento a concetti teorici. Si distinguono due tipi di relazioni:

- Le relazioni riflesse.
Sono relazioni in cui gli IE "riflettono" il concetto latente, quindi è il concetto latente che influisce sugli IE. Si può dire che, in questo caso, il concetto latente è la causa degli IE.
- Le relazioni formative.
Sono relazioni in cui gli IE "formano" il concetto latente, quindi gli IE influiscono sulla costruzione del concetto latente. In questo caso sono invece gli IE ad essere la causa del concetto latente.

Nel primo caso è chiaro che se gli IE sono causa del concetto latente essi esprimeranno concetti simili tra loro, quindi la relazione tra di esse sarà molto forte.

Nel secondo caso gli IE, formando il concetto latente, non hanno relazioni strette tra loro, anzi, più gli IE sono “slegati” tra loro più è probabile che essi spieghino meglio il concetto latente.

Ovviamente l’idea di questo tipo di relazione, si applica anche tra vari concetti latenti, quindi nella costruzione degli indicatori compositi. Nei capitoli successivi si vedrà come alcuni modelli statistici danno la possibilità di tradurre queste relazioni all’interno del modello stesso, generando quindi delle relazioni matematiche che non tralascino questo aspetto fondamentale.

2.3 Dagli Indicatori Compositi Data Driven agli Indicatori Compositi Model Based

La costruzione di un IC implica la ricerca di un’adeguata sintesi di un certo numero di VM al fine di ottenere una rappresentazione semplice del fenomeno multidimensionale. Di conseguenza, un IC può essere considerato come un Variabile Latente (VL), non direttamente misurabile, la cui stima può essere ottenuta attraverso i valori degli IE. In letteratura si possono riscontrare due diverse scuole di pensiero: Coloro che scelgono di aggregare le variabili in un singolo IC e coloro che non lo fanno e preferiscono utilizzare una serie di indicatori. Per la costruzione di un IC, in letteratura, sono stati proposti tre diversi approcci [190]:

- *Theory Based*, l’IC è ottenuto attraverso la combinazione di alcune variabili mediante una funzione specificata, suggerita da una teoria o da una buona conoscenza del fenomeno da analizzare;
- *Data Driven*, l’IC è ottenuto attraverso una sintesi ottimale delle variabili selezionate, che rappresentano le diverse facce del fenomeno analizzato;
- *Model Based*, l’IC è ottenuto dalla stima di un modello a più equazioni, descrivendo la relazione tra variabili osservate e con uno o più costrutti latenti da misurare, ma anche i rapporti tra loro.

2.3.1 Indicatori Compositi *Theory Based* e *Data Driven*

Gli *IC Theory Based* sono calcolati con formule semplici che combinano, di solito, poche variabili osservate. Questo approccio richiede una forte conoscenza del fenomeno oggetto di studio. Differentemente, un *IC Data Driven*, supera la mancanza di conoscenza del fenomeno. L'assenza di una conoscenza preliminare o di una teoria consolidata, causa spesso l'uso di un approccio guidato dai dati. Dunque il *Data Driven* È un approccio esplorativo che rientra in uno dei cinque principi fondamentali di Benzecrì su cui si basa l'analisi dei dati [9], secondo cui "i modelli devono seguire i dati e non viceversa". Così facendo, la dichiarazione di Benzecrì è invertita nel senso che sono i dati a seguire il modello al fine di costruire non solo IC descrittivi, ma che siano di aiuto per l'interpretazione e le decisioni di supporto. Diversi metodi sono stati proposti in letteratura [156] per la costruzione di un *IC Data Driven*, come ad esempio le tecniche di aggregazione, l'analisi di regressione lineare multipla, l'analisi delle componenti principali e l'analisi fattoriale, l'Alpha di Cronbach e la neutralizzazione dell'effetto di correlazione.

Tecniche di aggregazione. Prima di elaborare un IC è necessaria una trasformazione per omogeneizzare i vari IE; si definisce un sistema appropriato di pesi su cui si basa il calcolo di un IC, che inizia dal più semplice fino ad arrivare al più complesso. Ad esempio, l'Indice delle tecnologie dell'informazione e delle comunicazioni si basa sul metodo di aggregazione più semplice: comporta la classificazione dei paesi per ogni IE e quindi la somma delle classifiche del paese. L'indice di sostenibilità ambientale si basa sui punteggi standardizzati per ciascun indicatore che corrispondono alla differenza dell'indicatore per ciascun paese e la media dell'UE, diviso l'errore standard.

Analisi della Regressione Lineare Multipla. Si utilizza per calcolare la correlazione tra i coefficienti degli IE. Questo tipo di analisi può fornire informazioni circa la relazione tra un gran numero di indicatori X_1, X_2, \dots, X_n e il singolo indicatore di output Y . Un modello di regressione multipla è costruito sui coefficienti di regressione calcolati che sono i pesi relativi degli IE. Questo approccio, ad esempio, viene utilizzato per costruire l'indice nazionale di capacità di innovazione.

Analisi in Componenti Principali. Alcune applicazioni dell'Analisi in Componenti Principali (ACP) relative allo sviluppo di IC sono ad esempio:

- identificare la dimensione del fenomeno (*Environmental Sustainability Index*);
- raggruppare gli indicatori (*General Indicator of Science & Technology*);
- definire i pesi (*Internal Market Index*).

Il metodo dell'ACP è stato ampiamente usato per la costruzione di IC a partire da un gran numero di IE, sulla base proprio della correlazione tra essi (*Internal Market Index, Science and Technology Indicator*). In questi casi, le componenti principali, sono state utilizzate con l'obiettivo di combinare tanti IE in un unico IC per riflettere la massima percentuale possibile della variabilità totale presente nei dati. La prima componente principale dovrebbe di solito catturare una variabilità sufficiente per una adeguata rappresentazione dei dati originali (*Business Climate Indicator*). Tuttavia, in diversi casi, la prima componente principale da sola non arriva a spiegare più dell'80% della varianza totale degli IE e, per questo, deve essere combinata ad altre componenti principali per creare un singolo IC (*Success of software process implementation, Internal Market Index*).

Alpha di Cronbach. Un altro modo per studiare il grado di correlazione tra un insieme di IE è quello di utilizzare un coefficiente di affidabilità (o consistenza) denominato Alpha di Cronbach α . Questo coefficiente misura quanto bene un insieme di variabili (o indicatori) misura lo stesso costrutto sottostante. Un coefficiente di $\alpha = 0.80$ o superiore è considerato, nella maggior parte delle applicazioni, come prova che gli indicatori stanno misurando lo stesso costrutto sottostante. L'Alpha di Cronbach è stata considerata ad esempio per misurare l'*Index of Success of software process improvement*.

Neutralizzazione dell'effetto della Correlazione. Questo metodo fu applicato per l'aggregazione di tre IE in un IC per misurare l'intensità relativa dei problemi regionali delle Comunità da parte della Comunità Europea nel 1984. Gli IE misurano: a) PIL per impiegato in ECU, b) PIL pro capite PPS, e c) tasso di disoccupazione. Si basa su una forte correlazione tra IE, stimando un IC come media di IE rispetto alla correlazione.

L'approccio *Data Driven* usato in letteratura presenta diversi limiti rispetto al numero di IE usati, rispetto alla scelta del sistema di peso usato

per aggregare gli IE e rispetto all'assenza di relazioni tra gli IE e l'IC. Infatti, si suppone che:

- gli IE sono solo su scala numerica, non vengono utilizzati dati ordinali e nominali con perdita di preziose informazioni;
- gli IE assumono lo stesso ruolo senza distinguere tra variabili di input, di output e di outcome. Lo stesso vale per le variabili di mediazione e di moderazione il cui utilizzo potrebbe migliorare le informazioni;
- non viene menzionata esplicitamente la relazione tra IE e il loro IC (modello di misurazione riflessivo o formativo);
- non è consentito l'uso predittivo: il loro campo di applicazione è essenzialmente descrittivo con un uso limitato nel processo decisionale.

Inoltre, nella loro costruzione non viene considerata alcuna visione sistemica e non vengono prese in considerazione relazioni con altri IC. Al fine di superare queste restrizioni è stato proposto in letteratura un nuovo modo di costruire un IC che si muove da un approccio *Data Driven* ad uno *Model Based*.

2.3.2 Indicatori Compositi Model Based

Tra gli IC *Theory Based* e gli IC *Data Driven*, l'approccio *Model Based* consente di prendere in considerazione alcune informazioni *a priori* circa il fenomeno oggetto di studio, considerando la relazione tra l'IC di output e gli altri IC. Nell'approccio *Model Based*, un IC può prendere in considerazione conoscenze *a priori* del campo di interesse i) specificando il modello di misura (riflessivo, formativo o entrambi (MIMIC)); ii) definendo il ruolo degli IE nel modello; iii) contestualizzando l'IC rispetto agli altri IC secondo un determinato *path diagram*; iv) includendo ogni tipo di relazione tra gli IC (logica, gerarchica, temporale o spaziale). Per calcolare un IC *Model Based*, tenendo conto di tutte le informazioni *a priori*, il modello ad equazioni strutturali (*Structural Equation Model* - SEM) gioca un ruolo rilevante, in quanto, il calcolo dei pesi così come il processo di aggregazione degli IC non sono soggettivi. Entrambi i passaggi si basano su relazioni statistiche tra gli indicatori.

Il modello SEM [94] rappresenta un'estensione del modello lineare generale che stima simultaneamente le relazioni tra più VL indipendenti e tra

queste e la VL dipendente. Secondo questa metodologia, è possibile definire un IC come una VL multidimensionale non misurabile direttamente e legata alle sue VM secondo una relazione riflessiva, formativa o entrambe (questo definisce il modello di misura o anche detto modello esterno). Ogni IC è correlato agli altri IC attraverso equazioni di regressione lineare specificando il Modello Strutturale (o anche definito Modello Interno). Si ottiene così un “*Systemic*”, dove l’aggettivo “*Systemic*” deriva da una definizione data da Ludwig von Bertalanffy [11], secondo cui “un sistema è un insieme di elementi che interagiscono tra di loro, non solo una semplice aggregazione di IE, ma una serie di indicatori correlati ad altri indicatori mediante relazioni reciproche, espresse attraverso legami funzionali, sintetizzate in uno specifico modello”.

La scelta di utilizzare un modello SEM è particolarmente utile per diverse ragioni, in modo specifico:

- la possibilità di ottenere, simultaneamente e coerentemente al metodo di stima, una classifica di individui per ogni indicatore;
- la possibilità di confrontare gli indicatori nel tempo e nello spazio;
- la possibilità di stimare le relazioni ipotizzate senza fare ipotesi sulla distribuzione;
- la possibilità di definire un sistema ottimale di pesi;
- la possibilità di lavorare con un gran numero di variabili e poche osservazioni;
- la possibilità di valutare modelli complessi senza problemi di identificazione del modello;
- la possibilità di lavorare con dati mancanti e in presenza di multicollinearità.

Esistono due differenti approcci per stimare i parametri di un modello SEM: le tecniche *covariance-based* [86]; [87] e le tecniche *component-based* [208]; [204]; [107].

Il primo approccio viene utilizzato principalmente per confermare (o rigettare) delle teorie (ad esempio testare empiricamente un insieme di relazioni sistematiche tra più variabili), determinando quanto bene il modello teorico proposto può stimare la matrice di covarianza per un campione di dati. Invece, nelle tecniche *component-based*, la stima della VL gioca un ruolo

principale. Infatti, l'obiettivo di questi metodi è di fornire una stima delle VL in modo tale che esse siano fortemente correlate alle altre VL (secondo la struttura del *path diagram*) e le più rappresentative del proprio blocco di VM. Tra i diversi metodi che sono stati sviluppati per stimare un modello SEM, in questa tesi ci concentreremo sulle tecniche *component-based*, in modo particolare sul metodo del **PLS Path Modeling approach (PLS-PM)** [205]; [187], in quanto, la stima del IC gioca un ruolo chiave in questo processo di stima. Nel prossimo Capitolo, il PLS-PM sarà descritto in dettaglio sottolineando le sue principali proprietà e i vantaggi di utilizzare questo approccio per la stima di un sistema di IC.

3. PARTIAL LEAST SQUARES PATH MODELING

3.1 Introduzione

L'approccio *Partial Least Squares* (PLS) a *Structural Equation Model* (SEM), noto anche come *PLS-Path Modeling* (PLS-PM) è stato proposto come procedura di stima basata su componenti diversi dall'approccio LISREL basato sulla covarianza classica. Herman Wold [206] ha per primo formalizzato l'idea dei minimi quadrati parziali (*Partial Least Squares*) nel suo articolo sull'analisi delle componenti principali. La prima presentazione dell'approccio PLS finalizzato al *Path Modeling* con variabili latenti è stato pubblicato da Wold nel 1975 [205]. In altre presentazioni sul PLS-PM, che comparivano nello stesso anno [203]; [207], Wold [202] fornisce una discussione sulla teoria e l'applicazione del *PLS Path-Modeling* in econometria. Le principali referenze sull'algoritmo PLS sono Wold 1982 [208] e Wold 1985 [204]. Varie considerazioni sull'approccio PLS a SEM con ulteriori sviluppi sono riportate in Chin [27] e in Tenenhaus et al. [187].

Wold ha confrontato il SEM-ML [84] definito come “modellizzazione dura” alla modellizzazione *soft* del PLS. I due approcci SEM sono stati confrontati in Jöreskog e Wold [89]. Il PLS-PM è considerato un approccio di modellizzazione *soft*, in cui non sono richieste assunzioni forti, rispetto alle distribuzioni, alla dimensione del campione e alle scale di misura. Il PLS-PM segue le notazioni e i simboli dei SEM, incluso l'utilizzo di un *Path Diagram* per rappresentare le relazioni tra le variabili latenti e tra ciascuna variabile manifesta e la corrispondente variabile latente. Vale a dire, le p variabili manifeste sono rappresentate da rettangoli o quadrati, mentre i cerchi rappresentano le q variabili latenti. Le frecce definiscono le relazioni tra le variabili latenti e/o le variabili manifeste.

Come nei SEM, anche nel PLS-PM, le relazioni globali tra variabili manifeste e latenti sono modellizzate attraverso un sistema di equazioni. L'obiettivo del PLS-PM non è la riproduzione della matrice di covarianza campionaria, a differenza dell'approccio SEM. Per questo motivo, il PLS-PM è

considerato un approccio più esplorativo che confermativo [50]. Inoltre, il PLS-PM fornisce una stima diretta dei punteggi delle variabili latenti.

3.2 Il PLS-PM

Il PLS-PM è costituito da due elementi, il **modello di misura** (che si chiama anche **modello esterno**), che descrive le relazioni tra le variabili manifeste con le proprie latenti ed il **modello strutturale** (chiamato anche **modello interno**), che descrive le relazioni tra le variabili latenti. Entrambi i modelli sono descritti nelle sezioni successive.

3.2.1 Modello di misura

Una Variabile Latente (VL) ξ è una variabile non osservabile (o un costrutto) indirettamente descritta da un blocco di variabili osservabili x_k che sono chiamate Variabili Manifeste (VM) o indicatori. Esistono tre modi per relazionare le VM alle loro VL:

- Il modo riflessivo (o verso l'esterno);
- Il modo formativo (o verso l'interno);
- Il modo MIMIC (una miscela dei modi riflessivo e formativo).

Modo riflessivo

Nel *modo riflessivo*, ogni VM riflette la VL corrispondente.

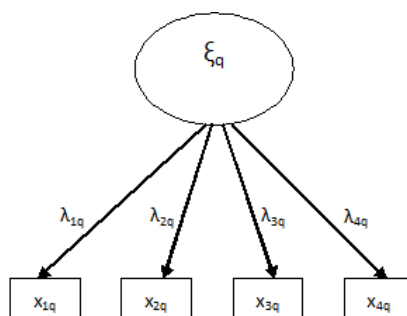


Fig. 3.1: Path-diagram di un modello riflessivo

Un blocco è definito *riflessivo* se la VL è considerata un fattore comune che si riflette nelle VM. Ciò implica che la relazione tra ogni VM x_{ij} (con i da 1 a q) e la corrispondente VL sia modellata come:

$$x_{pq} = \lambda_{pq}\xi_{pq} + \epsilon_q \quad (3.1)$$

dove ξ_{pq} è la VL esogena, λ_{pq} è il coefficiente di regressione semplice tra la VM e la VL chiamato *loading*.

Nel caso riflessivo, le VM dovrebbero essere fortemente correlate, poiché sono correlate con le VL di cui sono espressione. In altre parole, il blocco deve essere omogeneo. Esistono diversi strumenti per controllare l'omogeneità e l'unidimensionalità di un blocco riflessivo:

- Alpha di Cronbach;
- Rho di Dillon- Goldstein;
- L'analisi delle componenti principali del blocco.

Alpha di Cronbach: un blocco è considerato omogeneo se questo indice è maggiore di 0,7.

$$\alpha_q = \frac{\sum_{p \neq p'} \text{cor}(x_{pq}, x_{p'q})}{P_q + \sum_{p \neq p'} \text{cor}(x_{pq}, x_{p'q})} \times \frac{P_q}{P_q - 1} \quad (3.2)$$

dove P_q è il numero di VM nel blocco q e x_{pq} e $x_{p'q}$ sono due VM del blocco q –esimo. L'alpha di Cronbach è sensibile al numero di elementi della scala e generalmente tende a sottovalutare l'affidabilità interna della coerenza.

Rho di Dillon-Goldstein: misura l'affidabilità composita del blocco. Un blocco è considerato omogeneo se la sua affidabilità composita è maggiore di 0,7.

$$\rho_q = \frac{(\sum_{p=1}^{P_q} \lambda_{pq})^2}{(\sum_{p=1}^{P_q} \lambda_{pq})^2 + \sum_{p=1}^{P_q} (1 - \lambda_{pq}^2)} \quad (3.3)$$

Secondo Chin [27] il *Rho di Dillon-Goldstein* è considerato un indicatore migliore dell'omogeneità di un blocco rispetto all' *Alpha di Cronbach*.

L'analisi delle Componenti principali: un blocco è considerato omogeneo se, secondo la regola di Kaiser, il primo autovalore della matrice di correlazione è superiore a 1, mentre gli altri sono più piccoli [187].

Il primo indice presuppone che ogni VM sia ugualmente importante nella definizione della VL.

Nel *Rho di Dillon-Goldstein*, invece, questa assunzione non funziona perché si basa sui *loadings* del modello piuttosto che sulle correlazioni osservate tra le VM nel set di dati. Questo tipo di affidabilità tiene conto dei diversi *loadings* esterni delle variabili dell'indicatore. λ_{pq} rappresenta il *loading* esterno standardizzato della variabile indicatore i . L'affidabilità composta varia tra 0 e 1, dove i valori più alti indicano livelli di affidabilità più elevati. Viene generalmente interpretato nello stesso modo dell'Alpha di Cronbach. Tutte queste regole assumono, senza perdita di generalità, che le VL sono standardizzate e tutte le correlazioni tra le VM del blocco mostrano lo stesso segno. Quando questa ultima ipotesi non è verificata, è possibile solo indicare che il blocco è unidimensionale ma non omogeneo. Nel caso in cui l'ipotesi di unidimensionalità venga rifiutata, è possibile identificare alcuni gruppi di sotto blocchi unidimensionali considerando le correlazioni tra variabile e fattore visualizzate nel grafico dei *loadings*. Il PLS-PM è dunque una miscela di conoscenza a priori e analisi dei dati. Nel modo riflessivo, la conoscenza a priori riguarda l'unidimensionalità del blocco ed i segni dei *loadings* dei dati che devono adattarsi a questo modello. Se non lo fanno, possono essere modificati rimuovendo alcune VM che sono lontane dal modello. Un'altra soluzione è cambiare il modello e utilizzare il modo formativo.

Modo Formativo

Nel caso fromativo, le VL si suppongono essere generate dalle proprie VM.

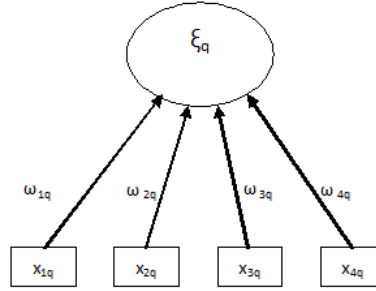


Fig. 3.2: Path-Diagram di un modello formativo

$$\xi_q = \sum_{p=1}^{P_q} \omega_{pq} x_{pq} + \delta_q \quad (3.4)$$

dove ω_{pq} è il coefficiente che collega ogni VM alla corrispondente VL e δ_q è l'errore che rappresenta la parte delle VL non spiegate dal blocco di VM.

L'assunto dietro questo modello è la seguente specifica predittiva:

$$E(\xi_q | x_{pq}) = \sum_{p=1}^{P_q} \omega_{pq} x_{pq} \quad (3.5)$$

il che implica che il vettore residuo sia $E(\delta_q) = 0$ e che esso non sia correlato con le VM.

Ogni VM o ogni set di VM rappresenta un livello diverso del concetto latente sottostante. Questo modello non assume l'omogeneità o l'unidimensionalità del blocco, per questo il blocco di VM può essere multidimensionale e gli indicatori non devono covariare. A differenza degli indicatori riflessivi, che sono sostanzialmente intercambiabili, non si prevedono correlazioni elevate tra gli elementi nei modelli di misurazione formativa. Infatti, l'elevata correlazione tra due indicatori formativi, nota anche come **collinearità**, può risultare problematica da un punto di vista metodologico e interpretativo. Quando sono coinvolti più di due indicatori, questa situazione viene chiamata *multicollinearità*. La collinearità può verificarsi perché

lo stesso indicatore viene immesso due volte o perché un indicatore è una combinazione lineare di un altro indicatore. Alti livelli di collinearità tra gli indicatori formativi sono una questione cruciale in quanto hanno un impatto sulla stima dei pesi e della loro significatività statistica, in particolare aumenta gli errori standard e riduce quindi la capacità di dimostrare che i pesi stimati sono significativamente diversi da zero e l'elevata collinearità può determinare che i pesi siano misurati in modo errato, così come nei loro segni invertiti. Per valutare il livello di collinearità, i ricercatori devono calcolare la *tolleranza*. La *tolleranza* rappresenta la quantità di varianza di un indicatore formativo non spiegato dagli altri indicatori nello stesso blocco. Può essere ottenuto in due fasi:

- in un primo momento prendiamo il primo indicatore formativo x_1 e lo regrediamo su tutti gli altri indicatori dello stesso blocco e calcoliamo la sua percentuale di varianza associata agli altri indicatori ($R_{x_1}^2$);
- quindi si calcola la tolleranza di questo indicatore (TOL_{x_1}):

$$TOL_{x_1} = 1 - R_{x_1}^2 \quad (3.6)$$

Una misura correlata di collinearità è il *Variance Inflation Factor (VIF)*, definito come reciproco della tolleranza:

$$VIF = \frac{1}{TOL_{x_1}} \quad (3.7)$$

Un valore di tolleranza di 0,20 o inferiore e un valore VIF di 5 o più indicano rispettivamente un potenziale problema di collinearità [64]. Se il livello di collinearità è molto elevato, si dovrebbe considerare la rimozione di uno dei corrispondenti indicatori [65].

Modo MIMIC

Il modo MIMIC è un *mix* dei modi riflessivo e formativo. I punteggi delle VL $\hat{\xi}_q$ standard associati alla q -esima VL ξ_q sono calcolati come una combinazione lineare del proprio blocco di VM tramite la relazione di pesi definita come:

$$\hat{\xi}_q = \sum_{p=1}^{P_q} \omega_{pq} x_{pq} \quad (3.8)$$

dove le variabili x_{pq} sono centrate e ω_{pq} sono i pesi esterni.

3.2.2 Modello strutturale

Nel PLS-PM, il modello strutturale specifica le relazioni tra le VL; una VL viene chiamata endogena se si suppone dipenda da altre VL altrimenti si parla di VL esogene. Nel modello strutturale ogni VL endogena è collegata ad altre VL mediante il seguente modello di regressione multipla:

$$\xi_j = \sum_{(q:\xi_q \rightarrow \xi_j)} \beta_{qj} \xi_q + \zeta_j \quad (3.9)$$

dove ξ_j è una VL endogena, β_{qj} è il coefficiente che collega la q - esima VL esogena alla j - esima VL endogena, esprimendo l'impatto delle VL esogene ξ_q sulla VL endogena collegata e ζ_j è l'errore nella relazione interna.

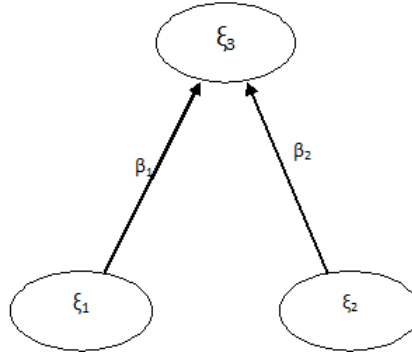


Fig. 3.3: Path-Diagram del modello strutturale

L'unica ipotesi di questo modello è quella che Wold ha chiamato *ipotesi predittiva* [208]: il vettore residuo ζ_j ha media zero e non è correlato con il predittore.

3.3 L'algoritmo dei minimi quadrati parziali

L'approccio PLS-PM [208];[204];[187] ai SEM è costituito da un algoritmo iterativo che calcola la stima delle VL, misurate da un insieme di VM e dalle relazioni tra di loro, mediante un sistema interdependente di equazioni basato su regressione multipla e semplice. L'idea è di determinare i punteggi delle VL attraverso un processo che, in modo iterativo, calcola prima una stima esterna e poi una stima interna.

L'algoritmo alterna la stima esterna con la stima interna. Esegue la stima

delle VL separatamente per ogni blocco e quindi aggiorna la stima con una stima interna. Quindi, nella fase di *stima esterna* l'algoritmo calcola i pesi w_{pq} , in base alla relazione tra VL e VM, dove q rappresenta il blocco latente q -esimo ad ogni VM per la stima della VL. L'algoritmo esegue la stima dei LV separatamente per ogni blocco e quindi aggiorna la stima delle VL, per la stima interna.

In particolare, gli algoritmi PLS comprendono tre fasi:

- Approssimazione Iterativa delle VL;
- Stima dei punteggi delle VL;
- Stima dei coefficienti dei *path*.

3.3.1 Prima fase: Approssimazione iterativa delle variabili latenti

La prima fase dell'algoritmo consiste di quattro passaggi [187]:

- Passo 1: Scegliere pesi esterni arbitrari iniziali;
- Passo 2: Calcolare l'approssimazione esterna delle VL e ottenere pesi interni;
- Passo 3: Calcolare l'approssimazione interna delle VL;
- Passo 4: Calcolare nuovi pesi esterni;
- Passo 5: Ripetere i passaggi dal 2° al 4° fino alla convergenza dei pesi esterni.

Passo 1: Scelta dei pesi esterni arbitrari iniziali

La procedura inizia scegliendo dei pesi arbitrari ω_{pq} (ad esempio tutti 1). Il processo iterativo assegnando qualsiasi combinazione lineare arbitraria non triviale di indicatori può servire come proxy esterno di una VL [70].

Passo 2: L'approssimazione esterna delle variabili latenti e dei pesi interni

In questo passaggio, le *proxy* esterne delle VL vengono calcolate come una combinazione lineare delle proprie VM centrate (stima esterna):

$$\nu_q = \sum_{p=1}^{P_q} w_{pq} x_{pq} \quad (3.10)$$

dove ν_q è la stima esterna standardizzata della q -esima VL ξ_q ; le x_{pq} sono VM centrate.

Nella stima del *modello interno* o *modello strutturale*, l'algoritmo aggiorna la stima delle VL, chiamate z_q , con il calcolo dei pesi interni $e_{qq'}$ (q' è una VL generica associata alla q -esima VL). Questi pesi vengono calcolati per ogni VL per riflettere quanto fortemente le altre VL sono connesse ad esso, considerando i collegamenti esistenti con le altre VL adiacenti di Q' :

$$z_q = \sum_{q'=1}^Q d_{qq'} e_{qq'} \nu_{q'} \quad (3.11)$$

$d_{qq'}$ è l'elemento generico della matrice quadrata D dell'ordine Q , dove $d_{qq'} = 1$ se la VL x_{iq} è connessa a $x_{i'q'}$ nel *path diagram* e $d_{qq'} = 0$ altrimenti.

I pesi interni $e_{qq'}$ sono calcolati in base a tre diverse alternative:

- *schema centroide* (lo schema originale di Wold), dove i pesi sono computati come:

$$e_{qq'} = \text{sign}[\text{cor}(v_q, v_{q'})] \quad (3.12)$$

Questa scelta mostra un inconveniente nel caso in cui la correlazione è di circa zero poiché il suo segno può cambiare per fluttuazioni molto piccole. Ma non sembra essere un problema nelle applicazioni pratiche.

- *schema fattoriale* (lo schema Lohmöller) dove i pesi sono calcolati come:

$$e_{qq'} = \text{cor}(v_q, v_{q'}) \quad (3.13)$$

Rispetto al metodo precedente, lo schema fattoriale è suggerito in tutti i casi in cui le correlazioni tra le VL sono più deboli.

- *schema dei pesi*, o schema strutturale, dove le VL collegate a ξ_q sono suddivise in due gruppi:

se $v_{q'}$ predice v_q

$$e_{qq'} = \text{cor}(v_q, v_{q'}) \quad (3.14)$$

oppure se $v_{q'}$ è predetto da v_q

$$e_{qq'} = \text{Coefficiente di regressione} \quad (3.15)$$

Passo 3: L'approssimazione interna delle variabili latenti

Le *proxy* interne delle VL sono calcolate come combinazioni lineari delle *proxy* esterne delle rispettive VL adiacenti, utilizzando i pesi interni precedentemente determinati.

Passo 4: La stima di nuovi pesi esterni

Una volta ottenuta una prima stima interna delle VL, l'algoritmo continua aggiornando i pesi esterni ω_{pq} . La stima dei pesi esterni dipende dal modello scelto. Esistono due modi per stimare questi pesi: *Modo A* e *Modo B*.

- *Modo A*: ogni peso esterno ω_{pq} è il coefficiente di regressione nella semplice regressione della p -esima VM del blocco q -esimo (x_{pq}) sulla stima interna z_q della q -esima VL. Infatti, dal momento che z_{pq} è standardizzata, i pesi esterni generici ω_{pq} sono ottenuti come

$$\omega_{pq} = \text{cov}(x_{pq}, z_q) \quad (3.16)$$

In questo caso, la VL si riflette sulle sue VM. Nel *path diagram* le frecce partono dalla VL verso le VM.

- *Modo B*: il vettore ω_q dei pesi ω_{pq} associati alle VM del blocco q è il vettore del coefficiente di regressione nella regressione multipla della stima interna z_q del valore q -esimo VL delle VM X_q

$$\omega_q = (X_q' X_q)^{-1} X_q' z_q \quad (3.17)$$

In questo caso, il concetto latente è formato dalle sue VM. Nel *path diagram* le frecce vanno dalle VM alle VL.

il PLS-PM con *Modo A* tende ad ottimizzare un criterio di covarianza [183], mentre il PLS-PM con *Modo B* ottimizza un criterio di correlazione [68]. Una piccola modifica dell'algoritmo PLS è necessaria per massimizzare un criterio di covarianza, ma la simulazione mostra che entrambi gli approcci sono in stretta corrispondenza [183]. La scelta di una certa modalità è soggetta a ragionamenti statistici e teorici e dipende tipicamente dalla decisione di definire un modello esterno come riflessivo o formativo. In particolare la modalità è strettamente legata alla natura del modello. Per un modello *riflessivo*, il *Modo A* è più appropriato, mentre il *Modo B* è

più adatto al modello *formativo*. Inoltre, il *Modo A* è suggerito per VL endogene, mentre *Modo B* per quelle esogene. Il *Modo A* e il *Modo B* possono essere utilizzati contemporaneamente quando il modello di misura è quello MIMIC.

Un PLS-PM generale sembra non ottimizzare alcun criterio, poiché Kramer ha dimostrato che il *Modo A* dell'algoritmo di Wold non si basa su equazioni stazionarie legate all'ottimizzazione di una funzione differenziabile due volte. Tuttavia, nel 2011 Tenenhaus [183] ha leggermente modificato il *Modo A* in cui un vincolo di normalizzazione viene posto sui pesi esterni piuttosto che sui punteggi delle VL. In particolare, è stato mostrato che la procedura di Wold, applicata a un PLS-PM dove il *nuovo modo A* viene utilizzato in tutti i blocchi, converge con il criterio:

$$\operatorname{argmax}_{\|\omega_q=1\|} \sum_{q \neq q'} c_{qq'} \operatorname{cov}^2(\mathbf{X}_q \omega_q, \mathbf{X}_{q'} \omega_{q'}) \quad (3.18)$$

quando viene utilizzato lo schema fattoriale per la stima interna delle VL.

In un approccio completamente basato sui dati, un'altra alternativa per aggiornare i pesi esterni è il *Modo PLS* [50]; [49]. In questa modalità ω_q è il vettore dei coefficienti di regressione in una regressione PLS di z_q su \mathbf{X}_q . Se l'algoritmo PLS-PM converge su una singola componente PLS-R, allora i pesi del *Modo PLS* saranno uguali al peso del *Modo A*: i dati sono espressione definitiva di un modello riflessivo. Se l'algoritmo PLS-PM converge su più componenti PLS-R, i dati vengono interpretati in un modello formativo: ogni sotto-blocco di VM rappresenta dimensioni diverse del concetto sottostante alle VL.

Questi tre passaggi vengono ripetuti finché la modifica dei pesi esterni tra due iterazioni non scende sotto un limite predefinito.

Passo 5: L'algoritmo di convergenza

La convergenza dell'algoritmo PLS-PM iterativo viene verificata secondo una regola di arresto, più spesso definita come:

$$\max |\omega_{pq}^{(s)} - \omega_{pq}^{(s-1)}| < 10^{-5} \quad (3.19)$$

dove s si riferisce alla s -esima iterazione.

3.3.2 Seconda fase: Stima dei punteggi delle VL

Una volta ottenuti i pesi finali ω_{pq} , i punteggi delle VL vengono calcolati come aggregati ponderati normalizzati delle VM:

$$\hat{\xi}_q \propto X_q \omega_q \quad (3.20)$$

3.3.3 Terza fase: Stima dei path coefficient

Nell'ultima fase dell'algoritmo del PLS-PM i *path coefficient* vengono stimati attraverso le regressioni multiple OLS tra i punteggi stimati delle VL, secondo la struttura del *path diagram*. Denotando ξ_j il generico vettore endogeno della VL e $\hat{\Xi}_{\rightarrow j}$ la matrice dei predittori latenti corrispondenti, il vettore del *path coefficient* per ciascuno ξ_j è:

$$\hat{\beta}_j = (\hat{\Xi}'_{\rightarrow j} \hat{\Xi}_{\rightarrow j})^{-1} \hat{\Xi}'_{\rightarrow j} \hat{\xi}_q \quad (3.21)$$

Nel caso di multicollinearità tra i punteggi stimati delle VL, al fine di ridurre la variabilità di stima, è possibile utilizzare la regressione PLS anziché la regressione OLS [50].

Dunque, in sintesi, la stima PLS-PM procede secondo il seguente schema iterativo:

Algoritmo PLS-PM

Inizializza l'algoritmo con la matrice X di VM grezze

Step1: Calcola un primo vettore casuale di pesi w_{pq}

repeat

Step2: Calcola la prima stima delle VL

for (q in 1:Q)

$$v_q = \sum_{p=1}^{P_q} w_{pq} x_{pq}$$

endfor

Step3: Aggiorna la precedente stima delle VL

for (q in 1:Q)

$$z_q = \sum_{q=1}^Q e_{qq}' v_q$$

endfor

Step4: Aggiorna la stima dei prsi w_{pq}

for (q in 1:Q)

for (p in 1: P_q)

$$w_{pq} = \text{cov}(x_{pq}, z_q)$$

```

         $w_q = (X_q' X_q)^{-1} X_q' z_q$ 
    endfor
endfor
Controlla la convergenza
    Fino a  $\sum |w_{pq}^{old} - w_{pq}^{new}| < \epsilon$ 

```

La convergenza¹ dell'algoritmo si ottiene se la somma della differenza assoluta dei pesi delle due stime successive esterne è minore di ϵ (piccolo valore reale positivo). Infine la stima interna dei *path coefficient* e dei *loadings* tra le VL, secondo la presunta relazione tra di essi, viene calcolata dal classico OLS per regressioni multiple/singole.

Dal punto di vista inferenziale, il PLS-PM non fa riferimento all'ipotesi di distribuzione su dati che utilizzano strumenti basati su inferenza computazionale, come le tecniche di ricampionamento. In particolare viene utilizzata la tecnica *Bootstrap* basata sull'estrazione con sostituzione di m campioni di dimensioni n (n è la dimensione originale del campione). Il modello viene stimato per ciascun campione di *Bootstrap* m -esimo per ottenere una distribuzione empirica dei parametri (pesi, *path coefficient* e *loadings*) e per calcolare un intervallo di confidenza appropriato. Questa procedura è fatta per i parametri sia del modello esterno (pesi e *loadings*) sia del modello interno (*path coefficient*). Intervalli che comprendono lo zero suggeriscono di eliminare le VM o le VL dal modello. Per confrontare i parametri stimati e la media delle ripetizioni di *bootstrap*, un rapporto tra la loro deviazione e la deviazione standard della distribuzione viene calcolato come una statistica di test classica.

3.4 Validazione del modello

La stima del modello fornisce misure empiriche dei modelli di misura (relazioni tra gli indicatori e i costrutti), nonché dei modelli strutturali (relazioni tra i costrutti). Le misure empiriche ci consentono di confrontare con la realtà i modelli di misurazione e modelli strutturali stabiliti in teoria, come rappresentati dai dati di esempio. In altre parole, possiamo determinare in che misura la teoria si adatta ai dati.

Più precisamente, la valutazione del modello di misura e modello strutturale del PLS-PM si basa su una serie di criteri di valutazione non parame-

¹ La convergenza dell'algoritmo PLS-PM è dimostrata per due blocchi. In caso di più blocchi la convergenza viene dimostrata solo empiricamente.

trici e utilizza procedure come il *Bootstrap* o il *Blindfolding*. Questo processo prevede una valutazione separata del modello di misura e del modello strutturale.

Inizialmente, la valutazione del modello si concentra sui modelli di misura. L'esame delle stime PLS-PM consente al ricercatore di valutare l'affidabilità e la validità delle misure costruttive.

Nel valutare i modelli di misurazione, dobbiamo distinguere tra costrutti riflessivi e formativi. I due approcci si basano su concetti diversi e pertanto richiedono l'esame di diverse misure valutative. I modelli di misura riflessivi vengono valutati sulla loro affidabilità e validità interna. Le misure specifiche includono l'affidabilità composita (come mezzo per valutare l'affidabilità interna di coerenza), la validità convergente e la validità discriminante. I criteri per i modelli di misurazione riflessivi non possono essere applicati universalmente a modelli di misurazione formativi. Con le misure formative, il primo passo è quello di garantire la validità del contenuto prima di raccogliere i dati e di stimare il PLS-PM. Dopo la stima del modello, vengono valutate le misure formative per la loro validità convergente, il significato, la pertinenza e la presenza di collinearità tra gli indicatori. Le stime del modello strutturale non vengono esaminate finché non sono state stabilite l'affidabilità e la validità dei costrutti. Se la valutazione dei modelli di misurazione riflessivi e formativi dimostra la qualità delle misure, vengono valutate le stime del modello strutturale. La valutazione PLS-PM del modello strutturale implica la capacità del modello di prevedere. Quindi, dopo l'affidabilità e la validità, i criteri di valutazione primari per i risultati PLS-PM sono i coefficienti di determinazione (valori di R^2), nonché il livello e l'importanza dei *path coefficient*. La valutazione degli esiti PLS-SEM può essere estesa ad analisi più avanzate (es. esaminando gli effetti mediatori e/o moderanti, considerando l'eterogeneità non osservata, i test a più gruppi, la varianza del metodo comune e altri).

3.4.1 Valutazione dei risultati dei modelli di misura riflessivi

La valutazione dei modelli di misurazione riflessivi comprende l'affidabilità composita per valutare la coerenza interna, l'affidabilità dei singoli indicatori e la varianza media estratta (Average Variance Extracted - AVE) per valutare la validità della convergenza. Inoltre, il criterio Fornell-Larcker e i *cross loadings* vengono utilizzati per valutare la validità discriminante. Per quanto riguarda i primi due, *consistenza interna* e *affidabilità degli indicatori individuali*, sono stati descritti in dettaglio sopra. La *validità della convergenza* è la misura in cui una stima corrisponde positi-

vamente a misure alternative dello stesso costrutto. Utilizzando il modello di campionamento di dominio, gli indicatori di un costrutto riflessivo vengono trattati con approcci diversi per misurare lo stesso costrutto. Pertanto, gli elementi che sono indicatori (misure) di un determinato costrutto devono convergere o condividere un'elevata percentuale di varianza.

Per stabilire la validità della convergenza, i ricercatori considerano i *loading* esterni degli indicatori, nonché l'AVE. Alti *loading* esterni su un costrutto indicano che gli indicatori associati hanno molto in comune, che vengono catturati dal costrutto. Questa caratteristica è comunemente chiamata "affidabilità dell'indicatore". Al minimo, i *loading* esterni di tutti gli indicatori dovrebbero essere statisticamente significativi.

Una misura comune per stabilire la validità convergente sul livello costruttivo è l'AVE [54] che esprime la parte della varianza del blocco spiegata da $\hat{\xi}_q$:

$$AVE_q = \frac{\sum_{p=1}^{P_q} \hat{\lambda}_{pq}^2}{\sum_{p=1}^{P_q} var(x_{pq})} \quad (3.22)$$

Questo criterio è definito come il valore medio dei *loading* quadrati degli indicatori associati al costrutto (cioè la somma dei *loading* quadrati divisi per il numero di indicatori). Un valore di AVE di 0,50 o superiore indica che, in media, il costrutto spiega più della metà della varianza dei suoi indicatori. Al contrario, un AVE di meno di 0,50 indica che in media più errori rimangono negli elementi rispetto alla varianza spiegata dal costrutto.

Pertanto, l'AVE è equivalente alla *comunalità* di un costrutto.

In un buon modello di misurazione, ogni VM è ben riassunta dalla propria VL. Quindi, per ogni blocco, un **indice di comunalità** viene calcolato come:

$$Com_q = \frac{1}{P_q} \sum_{p=1}^{P_q} cor^2(x_{pq}, \hat{\xi}_q) = \frac{1}{P_q} \sum_{p=1}^{P_q} \hat{\lambda}_{pq}^2 \quad (3.23)$$

cioè la media delle comunalità tra ogni VM del blocco q -esimo e $\hat{x}i_q$.

L'indice di comunalità misura la capacità delle VL di spiegare la varianza delle sue VM. Se lavoriamo su VM standardizzate, AVE e comunalità coincidono a meno della costante $1/P_q$

la Validità discriminante è la misura in cui un costrutto è veramente distinto da altri costrutti rispetto a standard empirici. Quindi, stabilire una validità discriminante implica che un costrutto sia unico e catturi fenomeni non rappresentati da altri costrutti nel modello. Sono state proposte misure alternative di validità discriminante. Un metodo per valutare la validità

discriminante è esaminando i *cross loadings* degli indicatori. In particolare, il *loading* esterno di un indicatore sul costrutto associato dovrebbe essere maggiore di tutti i suoi *loading* su altri costrutti (cioè i *cross loading*).

$$H_0 : cor(\xi_q, \xi_{q'}) = 1 \text{ control'ipotesi } H_1 : cor(\xi_q, \xi_{q'}) < 1 \quad (3.24)$$

La presenza di *cross loading* che superano i *loading* esterni degli indicatori rappresenta un problema di validità discriminante. Questo criterio è generalmente considerato piuttosto liberale in termini di determinazione della validità discriminante [64]. Cioè, è molto probabile indicare che due o più costrutti presentino una validità discriminante.

Il criterio Fornell-Larcker è un altro approccio che permette di valutare la validità discriminante. Esso confronta la radice quadrata dei valori AVE con le correlazioni di variabilità latente. In particolare, la radice quadrata dell'AVE di ogni costrutto dovrebbe essere maggiore della sua correlazione più alta con qualsiasi altro costrutto. La logica di questo metodo si basa sull'idea che un costrutto abbia più varianza con i suoi indicatori associati che con qualsiasi altro costrutto.

$$(AVE_q \text{ and } AVE_{q'}) > cor(\hat{\xi}_q, \hat{\xi}_{q'}) \quad (3.25)$$

Ciò significa che le VL spiegano meglio le proprie VM rispetto ad altre VL.

3.4.2 Valutazione dei risultati dei modelli di misura formativi

Molti ricercatori utilizzano erroneamente i criteri di valutazione del modello di misurazione riflessivo per valutare la qualità delle misure formative del PLS-PM, come dimostrato dalla revisione degli studi PLS-PM nelle discipline di gestione e marketing strategiche di Hair et al. [66].

I criteri di valutazione statistica per i modelli di misurazione riflessivi non possono essere trasferiti direttamente ai modelli di misurazione formativi in cui gli indicatori rappresentano probabilmente le cause indipendenti del costrutto e non necessariamente correlano altamente. I ricercatori dovrebbero concentrarsi sulla creazione della validità dei contenuti prima di valutare empiricamente i costrutti creati in modo formativo. Ciò richiede che gli indicatori formativi catturino tutte le facce (o almeno le più importanti) del costrutto. Nella creazione di costrutti formativi, i problemi relativi alla validità del contenuto sono affrontati nella definizione in cui

il ricercatore specifica chiaramente il dominio che gli indicatori sono destinati a misurare. I ricercatori devono includere una serie completa di indicatori che esaurisce il dominio del costrutto formativo. La mancata considerazione di tutte le facce del costrutto (cioè degli indicatori formativi rilevanti) comporta un'esclusione di parti importanti del costrutto stesso. La valutazione dei modelli di misurazione formativa richiede l'istituzione delle misure di validità convergente, la valutazione della collinearità degli indicatori e l'analisi dei relativi contributi assoluti, compresi il loro significato.

Validità convergente è il valore con cui una misura è correlata positivamente con altre misure (indicatori) dello stesso costrutto. Durante la valutazione dei modelli di misurazione formativa, dobbiamo verificare se il costrutto formativamente misurato sia fortemente correlato con una misura riflessiva dello stesso costrutto. Questo tipo di analisi è noto anche come **analisi della ridondanza** [26]. Il termine analisi della ridondanza deriva dalle informazioni del modello che sono ridondanti nel senso che è incluso nel costrutto formativo e anche in quello riflessivo.

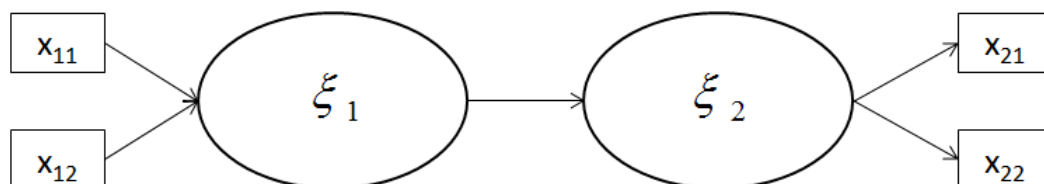


Fig. 3.4: Analisi della ridondanza per la validità convergente

La forza del *path coefficient* che collega i due costrutti è indicativa della validità del set designato di indicatori formativi nel formare il costrutto di interesse. Se l'analisi mostra una mancanza di validità convergente (per esempio il valore R^2 di $Y_{riflessivo} < 0,64$), gli indicatori formativi del costrutto $Y_{formativo}$ non contribuiscono ad un livello sufficiente. I costrutti formativi devono essere raffinati teoricamente/concettualmente mediante lo scambio e/o l'aggiunta di indicatori.

Per quanto riguarda il primo problema, la **collinearità tra gli indicatori**, è stata descritta in dettaglio nei paragrafi precedenti.

Significato e rilevanza degli indicatori formativi è un altro criterio im-

portante per valutare il contributo di un indicatore formativo. I valori dei pesi esterni possono essere confrontati l'uno con l'altro e possono pertanto essere utilizzati per determinare il relativo contributo di ciascun indicatore al costrutto o l'importanza relativa. Dobbiamo verificare se i pesi esterni nei modelli di misurazione formativa sono significativamente diversi da zero mediante la procedura di *bootstrap*. È importante notare che i valori dei pesi dell'indicatore formativo sono influenzati da altre relazioni del modello (si veda l'algoritmo PLS-PM). I pesi degli indicatori non significativi non dovrebbero essere automaticamente interpretati come indicativi di una scarsa qualità del modello di misura. Inoltre i ricercatori dovrebbero anche considerare **il contributo assoluto** di un indicatore formativo al suo costrutto, cioè l'informazione che un indicatore fornisce senza considerare altri indicatori. Il contributo assoluto è dato dal *loading* esterno dell'indicatore formativo, che è sempre fornito insieme ai pesi degli indicatori. A differenza dei pesi esterni, i *loading* esterni derivano da regressioni singole di ciascun indicatore sul relativo costrutto. Quando il peso esterno di un indicatore non è significativo, ma il suo *loading* esterno è elevato (cioè superiore a 0,50), l'indicatore deve essere interpretato come assolutamente importante ma non come relativamente importante. In questa situazione, l'indicatore sarebbe generalmente mantenuto. Ma quando un indicatore ha un peso non significativo e il *loading* esterno è inferiore a 0,50, il ricercatore dovrebbe decidere se mantenere o eliminare l'indicatore esaminando se la relativa rilevanza teorica e il contenuto potenziale si sovrappongono ad altri indicatori dello stesso costrutto.

3.4.3 Valutazione dei risultati dei modelli strutturali

Una volta confermato che le misure dei costrutti sono affidabili e valide, il passo successivo riguarda la valutazione dei risultati del modello strutturale. Questo richiede di esaminare le capacità predittive del modello e le relazioni tra i costrutti. I criteri chiave per la valutazione del modello strutturale PLS-SEM sono il significato dei *path coefficient*, del livello dei valori R^2 , della dimensione dell'effetto f^2 , della pertinenza predittiva Q^2 e la dimensione dell'effetto q^2 .

I **path coefficient del modello strutturale** rappresentano le relazioni ipotizzate tra i costrutti. Se un coefficiente è significativo in ultima analisi dipende dal suo errore standard che si ottiene mediante il *bootstrap*. L'errore standard del *bootstrap* di calcolare il *valore empirico*

$$t = \frac{p_{qj}}{se_{p_{qj}}^*} \quad (3.26)$$

quando il valore t empirico è maggiore del *valore critico*, il coefficiente è significativo ad una certa probabilità di errore (cioè livello di significatività); i valori critici comunemente usati per i test a due vie sono 1,65 (livello di significatività = 10%), 1,96 (livello di significatività = 5%) e 2,57 (livello di significatività = 1%). Invece dei valori t , i ricercatori riportano regolarmente i valori p che corrispondono alla probabilità di rifiutare erroneamente l'ipotesi nulla, considerando i dati a disposizione. Oltre al calcolo dei valori di t e p , è possibile determinare l'intervallo di confidenza per una probabilità di errore predefinita.

Il coefficiente R^2 è una misura della precisione predittiva del modello e viene calcolata come la correlazione quadrata tra valori effettivi e predittivi di un determinato costrutto endogeno. Rappresenta la quantità di varianza nei costrutti endogeni spiegati da tutti i costrutti esogeni collegati ad esso. Il valore R^2 varia da 0 a 1 con livelli superiori che indicano livelli di precisione predittiva più alti; il valore R^2 accettabile dipende dalla complessità del modello e dalla disciplina di ricerca [66].

La dimensione dell'effetto f^2 è una misura aggiuntiva nella valutazione del valore R^2 di tutti i costrutti endogeni. La variazione di R^2 viene esplorata per vedere se una specifica VL esogena ha un impatto sostanziale sul R^2 :

$$f^2 = \frac{R_{included}^2 - R_{excluded}^2}{1 - R_{included}^2} \quad (3.27)$$

dove $R_{included}^2$ e $R_{excluded}^2$ sono l' R^2 della VL endogena quando una determinata VL esogena è inclusa o esclusa del modello. Le linee guida per la valutazione del f^2 sono proposte da Cohen [39]:

- se $f^2 \approx 0.02 \rightarrow$ impatto basso
- se $f^2 \approx 0.15 \rightarrow$ impatto medio
- se $f^2 \approx 0.35 \rightarrow$ impatto alto

la Rilevanza predittiva Q^2 è un indicatore della rilevanza predittiva del modello sviluppato da Stone [178] e Geisser [57]. L'adattamento PLS-PM di questo approccio segue una procedura di *blindfolding*. Dato un blocco di n e p VM, la procedura esegue una parte del blocco considerato durante le

stime dei parametri e quindi tenta di stimare la parte omessa utilizzando i parametri stimati. Per stimare il modello, il valore omesso viene in genere sostituito con la media della variabile (sebbene si possano utilizzare altre tecniche di imputazione [27]). Sulla base del modello stimato, le stime per il valore omesso vengono confrontate con i valori osservati, utilizzando la differenza quadrata (E). Allo stesso tempo, anche la differenza tra i valori medi delle variabili (o valori altrimenti imputati) ed i valori osservati vengono confrontati utilizzando la differenza quadrata (O). Questa procedura viene ripetuta finchè ogni punto dati è stato omesso e stimato. La misura predittiva per queste VM è la seguente:

$$Q^2 = 1 - \frac{\sum_m E_m}{\sum_m O_m} \quad (3.28)$$

dove m è il numero di volte in cui viene ripetuta la procedura per assicurare che ogni punto dati sia omesso. Q^2 rappresenta una misura di come il valore ben osservato viene ricostruito dal modello e dai suoi parametri di stima [23]. Quando il PLS-PM mostra rilevanza predittiva, prevede accuratamente i punti dati degli indicatori nei modelli di misurazione riflessi nei costrutti endogeni di un singolo elemento (la procedura non si applica ai costrutti endogeni formativi). $Q^2 > 0$ implica che il modello abbia rilevanza predittiva mentre $Q^2 < 0$ rappresenta una mancanza di rilevanza predittiva.

Nel modello strutturale, valori di Q^2 maggiori di zero per una certa variabile latente endogena riflessa indicano la rilevanza predittiva del modello del path per questo particolare costrutto. Al contrario, i valori 0 o sottostanti indicano una mancanza di rilevanza predittiva. Simile all'approccio di dimensioni dell'effetto f^2 per la valutazione dei valori di R^2 , l'impatto relativo della pertinenza predittiva può essere confrontato mediante la misura alla dimensione dell'effetto q^2 , formalmente definita come segue:

$$q^2 = \frac{Q_{included}^2 - Q_{excluded}^2}{1 - Q_{included}^2} \quad (3.29)$$

dove $Q_{included}^2$ e $Q_{excluded}^2$ sono il valore di Q^2 della VL endogena quando una determinata VL esogena è inclusa o esclusa dal modello. Come misura relativa della pertinenza predittiva, i valori di 0,02, 0,15 e 0,35 indicano che un costrutto esogeno ha una rilevanza predittiva di piccole, medie o grandi dimensioni per un certo costrutto endogeno [66].

Diverse forme di Q^2 possono essere ottenute da procedure diverse per prevedere osservazioni formulate dal modello. Nella *comunalità convalidata*

Q^2 la previsione delle osservazioni viene effettuata dal computo calcolato e dai loading stimati. *Ridondanza Crossvalidata* Q^2 è ancora basata sui loading stimati ma il computo è previsto dal modello strutturale utilizzando i path coefficient stimati. il Q^2 basato sulla ridondanza è applicabile solo alle osservazioni di VM dei blocchi endogeni, mentre Q^2 basato sulla comunaltà può essere applicato a tutti le VM [23].

Tenenhaus et al. [185]; [187] ha proposto una soluzione PLS **Goodness-of-Fit (GoF)** come *una soluzione operativa a questo problema, che può essere inteso come un indice per convalidare il modello PLS globalmente*. Il GoF può essere calcolato come media geometrica della media della comunaltà e della media degli R^2 :

$$GoF = \sqrt{\overline{Com} \times \overline{R^2}} \quad (3.30)$$

dove $\overline{R^2} = \frac{1}{J} \sum_{j=1}^J R_j^2$.

Il GoF è un compromesso tra la qualità del modello esterno e la qualità del modello interno, quindi viene ottenuto un indice normalizzato riportando ogni parte al suo valore massimo. In particolare per la stima esterna (la prima parte della formula è la comunaltà media) per ogni blocco il massimo è il primo autovalore, poiché la prima componente principale spiega la variabilità massima, mentre per la stima interna il massimo è dato dalla prima correlazione canonica al quadrato. Per verificare la significatività del GoF, è possibile costruire un intervallo di confidenza con la tecnica *Bootstrap* come pure per R^2 .

Henseler e Sarstedt [74] criticano l'utilità del GoF sia concettualmente che empiricamente. La loro ricerca dimostra che il GoF non rappresenta un criterio di buona qualità per i PLS-SEM. Usando i dati simulati, hanno illustrato che il GoF non è adatto per la convalida del modello. Per alcuni tipi specifici di convalida del modello, tuttavia, l'applicazione del GoF ha senso. Cioè, quando si tratta di convalidare modelli che non differiscono nella loro struttura ma nei loro indicatori (riflessi), il GoF è la statistica da scegliere. Se il modello strutturale rimane costante, il GoF può valutare indirettamente i cambiamenti relativi alla validità della convergenza come espresso dall'AVE [54]. Il GoF è anche molto utile per i confronti dei dati (cioè variando i dati pur mantenendo costante il modello). Di conseguenza, il GoF è meglio applicato nei confronti di gruppi [168] e valutazioni di eterogeneità non osservate, come avviene per la procedura REBUS-PLS. In questi casi, il GoF può rispondere a domande su come i sottoinsiemi dei da-

ti possono essere spiegati da un particolare modello. Tuttavia, dato che il GoF non è applicabile anche ai modelli di misurazione formativi e non penalizza gli sforzi di over-parametrizzazione, i ricercatori non sono invitati a utilizzare questa misura. Per un blocco formativo, si potrebbe sostituire nella formula del GoF la comunaltà di blocco all' R^2 tra il *proxy* interno del blocco formativo e le VM del blocco. Un altro punto di partenza potrebbe essere valutare i pesi dei blocchi formativi. La ricerca futura dovrebbe formulare suggerimenti più concreti su come migliorare il GoF e dimostrarne i miglioramenti sia con il ragionamento concettuale sia con le simulazioni di Montecarlo [74].

3.5 L'importanza dei punteggi degli Indicatori Compositi

Una caratteristica fondamentale del metodo PLS-PM è l'estrazione dei punteggi per gli Indicatori Compositi (IC). Nel sistema di IC costruiti con il PLS-PM, è possibile ottenere i punteggi per ogni IC, esogeni ed endogeni, e per ogni IC è possibile fare una classifica tra le unità. Inoltre, il PLS-PM fornisce informazioni sulla relativa importanza dei costrutti nello spiegare altri costrutti nel modello strutturale. Le informazioni sull'importanza dei costrutti sono rilevanti per trarre conclusioni. Per l'analisi SWOT (*Strengths, Weaknesses, Opportunities, Threats*) è utile estendere i risultati degli esiti PLS-SEM di base utilizzando i punteggi delle variabili latenti [75]. I risultati del PLS-PM tengono conto delle prestazioni di ciascun costrutto. Inoltre, sono considerati i valori medi degli IC. Per uno specifico IC endogeno, la SWOT confronta gli effetti totali del modello strutturale (importanza) ed i valori medi dell'IC (performance). Di conseguenza, si possono trarre conclusioni su due dimensioni (vale a dire sia sull'importanza che sulla performance), particolarmente importanti per dare priorità alle azioni. L'analisi SWOT si basa su un grafico a dispersione in cui ogni IC è posizionato secondo la sua media e il suo *path coefficient* rispetto al IC *target*. L'asse x rappresenta gli effetti complessivi degli IC sull'IC *target* (cioè la loro importanza). L'asse y descrive i punteggi medi dei costrutti di queste VL (cioè la loro prestazione). In questo modo il grafico a dispersione è diviso in quattro aree o quadranti:

- la prima è l'area più critica, anche detta area di intervento immediato, perché gli IC in quest'area hanno un alto impatto, ma un basso valore medio;

- la seconda area è l'area di monitoraggio, in cui gli IC hanno un valore medio e un *path coefficient* basso;
- la terza, è l'area da migliorare perché gli IC hanno un alto valore medio e un coefficiente basso.
- la quarta è l'area da mantenere, in cui gli IC hanno un elevato valore per la media e per il *path coefficient*.

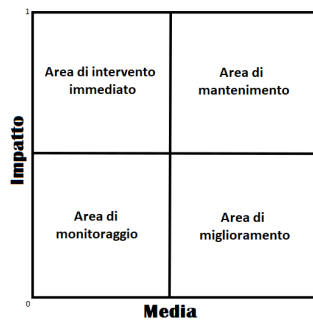


Fig. 3.5: Grafico Swot

3.6 Software disponibili per il PLS Path Modeling

Per lungo tempo **LVPLS** 1.8 [107] è stato l'unico software disponibile per il PLS Path Modeling. Il programma basato su DOS comprende due moduli diversi per la stima del PLS-PM. Il metodo **LVPLSC** analizza la matrice di covarianza delle variabili osservate, mentre il modulo **LVPLSX** è in grado di elaborare i dati grezzi. Per specificare il file di input è necessario un editor esterno. La specifica di input richiede che i parametri del programma siano definiti in posizioni specifiche nel file. I risultati sono riportati in un file di testo normale. Il programma offre *blindfolding* e *bootstrap* come metodi di ricampionamento nel caso in cui i dati grezzi siano stati analizzati. Nell'analizzare le matrici di covarianza/correlazione, non è possibile applicare le tecniche di ricampionamento [119].

Nel corso degli anni sono stati sviluppati altri software per il PLS-PM. L'elenco include **SmartPLS** [149], **XLSTAT-PLSPM** [51] in collaborazione con Addinsoft France, (<http://www.xlstat.com/en/products/xlstat-plspm/>) e il **plsmpm** package [164]. **SmartPLS** e **XLSTAT-PLSPM** sono sorgenti chiuse e **plsmpm** è concesso in licenza con la General Public License (GPL \geq 2).

- **SmartPLS:** **SmartPLS** è un software stand-alone specializzato per il PLS-PM. è costruito su una piattaforma Java Eclipse che lo rende indipendente dal sistema operativo. Il modello viene specificato tramite drag&drop trascendendo il modello strutturale per le variabili latenti e assegnando gli indicatori alle variabili latenti. è possibile caricare i file di dati di vari formati. Dopo aver creato un modello, i coefficienti vengono aggiunti al grafico. L'output più dettagliato è fornito in formato testo, LATEX e HTML. Il grafico che rappresenta il modello può essere esportato in PNG. Oltre ai metodi di *bootstrap* e *blindfolding*, supporta la specificazione degli effetti di interazione. Una caratteristica particolare di SmartPLS è la "*finite mixture routine*" (FIMIX), un metodo per affrontare l'eterogeneità non osservata. [167]; [168] .
- **XLSTAT-PLSPM:** **XLSTAT** [1] è un software statistico modulare basato su Microsoft Excel per l'inserimento dei dati e la visualizzazione dei risultati, ma i calcoli vengono eseguiti da componenti software autonome. **XLSTAT-PLSPM** è integrato in XLSTAT come modulo per la stima dei modelli PLS-PM. è stato sviluppato da un team di ricerca del Dipartimento di Matematica e Statistica dell'Università di Napoli in Italia e Addinsoft in Francia e implementa tutte le funzionalità metodologiche e le ultime risultanze della tecnologia PLEASURE (Partial LEAst Squares Struttural Relationship Estimation) di Esposito Vinzi et al. [51]. Le caratteristiche speciali di XLSTAT-PLSPM sono confronti multi-gruppo [28] e l'approccio di segmentazione REBUS [50] per il trattamento dell'eterogeneità non osservata.
- **plsmp in R:** il pacchetto **plsmp** implementa il metodo PLS con l'accento sui modelli di equazione strutturale in R. Il metodo di adattamento 'plsmp.fit' restituisce un elenco che include tutti i parametri stimati e quasi tutte le statistiche associate con i PLS-PM. Il metodo di summary fornisce una panoramica dei seguenti elementi: modello esterno, modello interno, VL scalate, pesi esterni, *loading*, matrice dei *path coefficient*, R^2 , correlazioni esterne, modello interno riassuntivo, effetti totali, unidimensionalità, bontà di misura, risultati *bootstrap* (solo se attivati) e la matrice di dati.

4. COSTRUTTI DI ALTO ORDINE NEL PLS-PM

4.1 Introduzione

Come è stato detto nei capitoli precedenti, molti fenomeni sono complessi e basati su diversi livelli di astrazione. Pertanto, il PLS-PM è uno strumento idoneo per l'indagine di questo tipo di modelli con un elevato livello di astrazione e nei casi in cui la costruzione di un sistema di indicatori compositi dipenda da diversi livelli di costruzione. Circa 25 anni fa Noonan e Wold [132] osservarono che: *"L'analisi dei path con VL strutturate gerarchicamente nell'ambito del PLS è in una fase iniziale di sviluppo e la ricerca è ancora in corso"*. Fortunatamente, negli ultimi anni, è stata intrapresa una ricerca sull'uso di modelli per costrutti di alto ordine con il PLS-PM e sono state sviluppate diverse applicazioni. L'uso di questi modelli ha consentito ai ricercatori di estendere l'applicazione del PLS-PM a modelli più avanzati e complessi. Nel contesto dei modelli PLS-PM, i costrutti di alto ordine hanno mostrato una popolarità crescente negli ultimi anni. Alcuni autori hanno discusso sia dei contributi teorici che empirici dei modelli gerarchici [46]; [81]; [83]; [111]; [198]; [38]; [34]; [35]; [37]. I modelli ad equazioni strutturali basate sulle componenti (CB-SEM) ed, in particolare, il PLS-PM possono essere utilizzati per stimare i parametri nei modelli di alto ordine [198]. Per i SEM basati sulla covarianza, sono generalmente disponibili linee guida e illustrazioni empiriche [46]. Per il PLS-PM, le linee guida sono principalmente disponibili per i modelli di alto ordine con relazioni riflesse ([108]; [198]; [208]). Tuttavia, Ringle et al. [83] mostrano che i modelli di Alto Ordine (AO) con relazioni riflesse nel primo e nel secondo ordine della gerarchia rappresentano solo una minoranza (20 %) dei modelli applicati in *MIS Quarterly*. Pertanto, esiste una grande necessità di orientamenti sull'utilizzo di modelli gerarchici con relazioni formative del PLS-PM, come dimostra chiaramente il modello del secondo ordine per il capitale sociale di Koka e Prescott [95]. Modelli per costrutti di ordine superiore, noti anche come *modelli gerarchici* o *costrutti multidimensionali* sono rappresentazioni esplicite di costrutti multidimensionali che esistono ad un livello superiore di astrazione e sono correlati ad altri costrutti a

un livello simile di astrazione e ne mediano l'influenza da o verso le loro dimensioni sottostanti [27]; [26]. Law et al. [99] definiscono “[...] un costrutto come multidimensionale quando consiste di un numero di attributi o dimensioni interconnesse ed esiste in domini multidimensionali. Queste dimensioni possono essere concettualizzate sotto un’astrazione complessiva ed è teoricamente significativo e parsimonioso usare questa astrazione complessiva come una rappresentazione delle dimensioni.” La creazione di modello con componenti di ordine più elevato, di solito nel contesto del PLS-PM [108], comporta spesso il compimento di due livelli di secondo ordine che contengono due strati di costrutti. Questo tipo di modello è spesso limitato ad una struttura gerarchica del secondo ordine e può essere definita come un costrutto che coinvolge più di una dimensione [46]; [81]; [98]; [111]; [130]; [135]. Come tale, si può distinguere dai costrutti unidimensionali, che sono caratterizzati da una singola dimensione sottostante [130].

Ci sono tre ragioni principali per includere i costrutti di alto ordine nei modelli PLS-PM.

- Innanzitutto, creando modelli per costrutti di alto ordine, i ricercatori possono ridurre il numero di relazioni nel modello strutturale, rendendo il PLS-PM più parsimonioso e più facile da cogliere.
- In secondo luogo, i modelli per costrutti di alto ordine sono importanti se i costrutti sono fortemente correlati; le stime dei rapporti del modello strutturale possono essere mal stimati a causa di problemi di collinearità e la validità discriminante non può essere stabilita. In situazioni caratterizzate dalla collinearità tra i costrutti, un costrutto di secondo ordine può ridurre tali problemi di collinearità e risolvere problemi di validità discriminante.
- In terzo luogo, la creazione di modelli per costrutti di ordine superiore può risultare utile anche se gli indicatori formativi presentano elevati livelli di collinearità. A condizione che la teoria supporti questo passo, i ricercatori possono dividere l'insieme di indicatori e stabilire costrutti separati in una struttura ad alto livello.

L'utilità di questi modelli si basa su un certo numero di motivi teorici ed empirici [46]. I sostenitori dell'utilizzo di costrutti di ordine superiore sostengono che tali modelli permettono una parsimonia più teorica e riducono la complessità del modello [46]; [99]; [111]. Edwards [46] riassume questo argomento come utilità teorica; la teoria richiede costrutti generali

costituiti da dimensioni specifiche. Ciò è strettamente correlato al compromesso tra l'accuratezza e la generalizzazione, come suggerito da Gorsuch [60], il quale sostiene che “i fattori riguardano aree restrittive di generalizzazione dove l'accuratezza è grande, mentre fattori di ordine superiore riducono la precisione per un aumento della larghezza della generalizzazione”. Anche Law et al. [99] affermano che “considerando le dimensioni come un insieme di variabili individuali esclude qualsiasi conclusione generale tra costruzione multidimensionale e altri costrutti”.

4.2 Stima dei costrutti nei modelli di ordine superiore

Edwards [46] ha proposto un quadro analitico integrativo sulla base dei SEM, che permette l'inclusione simulata di costrutti di ordine superiore e le loro dimensioni come VL. In un modello strutturale, i costrutti di ordine superiore possono servire come causa o effetto da incorporarsi in una rete nomologica. Questo approccio ci permette inoltre di trarre gli effetti (indiretti) dei costrutti di ordini inferiori, sui risultati del costrutto di ordine superiore come prodotto a coppie di *loading* (o *pesi* per i costrutti formativi) e coefficienti dei risultati. Inoltre, i SEM consentono la specificazione esplicita della direzione dei rapporti tra VM e VL [47].

4.2.1 Modelli Molecular and Molar per costrutti di alto ordine

A causa della natura deterministica dell'algoritmo PLS che pesa gli indicatori di misura per creare i punteggi dei costrutti, si possono modellare due tipi di costrutti ad alto ordine, che Chin e Gopal definiscono come *Molecular* e *Molar Higher-Order Constructs* [30]. In sostanza, questi due modelli possono essere distinti sulla base delle direzioni dei rapporti tra le VM e VL [98].

Per i costrutti di ordine superiore *Molecular* (o modelli di costrutti riflessivi), le VM sono influenzate dalle VL ($VL_j \rightarrow VM_i$), mentre per i costrutti di alto ordine *Molar* (o modelli di costrutti formativi), il rapporto è invertito ($VL_j \leftarrow VM_i$).

4.2.2 Tipi di modelli per costrutti di alto ordine

Ciascuno dei modelli di alto ordine è caratterizzato da differenti rapporti tra i costrutti di alto ordine e le VL: la relazione riflessiva e la relazione formativa. Come si può vedere nella figura 4.1, ci sono quattro tipi principali

di modello per costrutti di alto ordine discussi nella letteratura esistente [81]; [198] e utilizzati nelle applicazioni [83]. Questi tipi di modello hanno due elementi: il *costrutto di ordine superiore (High Order Construct - HOC)*, che cattura l'entità più astratta e i *costrutti di basso ordine (Low Order Construct - LOC)* che catturano le sottomisure dell'entità astratta.

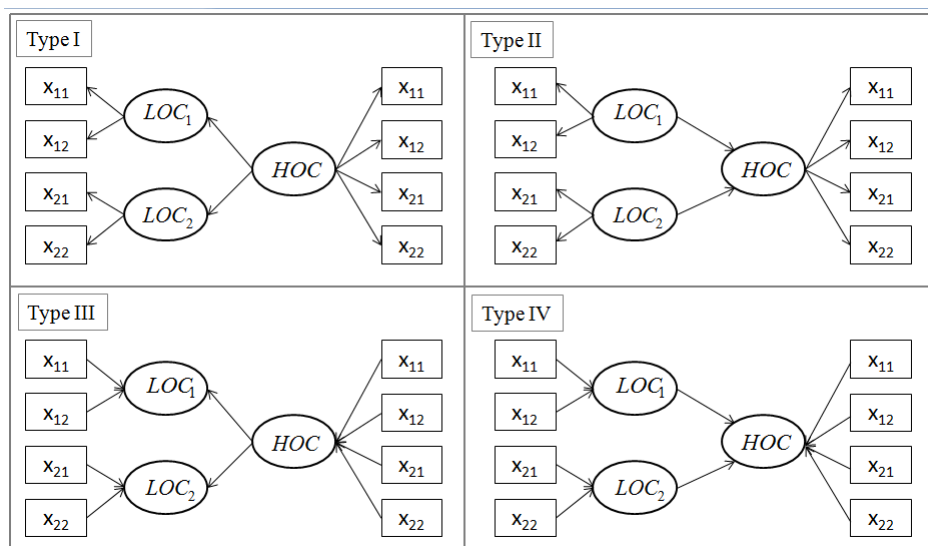


Fig. 4.1: Tipi di costrutti di alto ordine

- Uno dei modelli più frequentemente applicati nei SEM tra i ricercatori è oggi il modello di misura *Riflessivo-Riflessivo* conosciuto come *Costrutto di Secondo Ordine di Tipo I*.
- In secondo luogo, il modello di misura *Riflessivo-Formativo* o anche *costrutto di secondo ordine di Tipo II*. Questo modello comprende modelli di misurazione riflessivi ed è un modello strutturale formativo. Secondo il chiarimento di Chin, i LOC sono costruiti in modo selettivo, e non condividono una causa comune, ma piuttosto costituiscono un concetto generale che interpone interamente l'impatto sulle variabili endogene successive [27].
- In terzo luogo, il modello di misura *Formativo-Riflessivo* o anche *costrutto di secondo ordine di Tipo III*. In questo caso, ciascuna VM (in-

dicatore) andrà a formare i LOC e allo stesso tempo l'effetto causale dell'HOC verrà esercitato sui LOC che compongono l'indicatore.

- Infine, il modello di misura *Formativo-Formativo* è il meno utilizzato tra i modelli strutturali. Questa applicazione è appropriata quando sia gli HOC che i LOC sono costrutti formativi.

4.3 *Approcci basati sul PLS-PM per la stima di modelli di alto ordine*

Nell'ambito del PLS-PM, in letteratura sono presentati tre approcci principali per affrontare i modelli VL di alto ordine. Questi approcci saranno descritti dettagliatamente nei prossimi paragrafi.

4.3.1 *L'approccio ad indicatori ripetuti*

Il disegno originario di Wold per il PLS-PM non considera VL di ordine superiore; ogni costrutto deve essere necessariamente correlato ad un insieme di variabili osservate per essere stimato. Su questa base, Lohmöller [108] ha proposto una procedura per il caso di costrutti gerarchici, il cosiddetto *modello di costrutto gerarchico* [208] o *approccio ad indicatori ripetuti* [208]; [108] o *approccio super-block* [187], che è l'approccio più diffuso per la stima di costrutti di alto ordine tramite il PLS [195]; [199]; [212].

La procedura è molto semplice: *“un fattore di secondo ordine è direttamente misurato da variabili osservate per tutti i fattori del primo ordine. Questo approccio ripete il numero di VM utilizzate, così che il modello può essere stimato con l'algoritmo standard PLS ”* [142]. Le VM, misurando ciascuna VL di primo ordine, vengono semplicemente ripetute per rappresentare il costrutto ad alto ordine. Ad esempio, se una VL di secondo ordine è costituita da due VL di primo ordine sottostanti, ognuna con due VM, può essere specificata la VL di secondo ordine utilizzando tutte le VM dei sottostanti VL di primo ordine e quindi la VL di secondo ordine sarà formata da quattro VM.

Di conseguenza, le VM vengono utilizzate due volte: per le VL di primo ordine (*loading primario*) e per le VL di secondo ordine (*loading secondario*). Avendo così specificato il modello esterno (modello di misura), il modello interno (il modello strutturale) rappresenta la componente gerarchica del modello, in quanto rappresenta i *loading* della VL del secondo ordine sulle VL di primo ordine. Ovviamente, questo approccio può essere facilmente esteso ai modelli di alto ordine [132] [33].

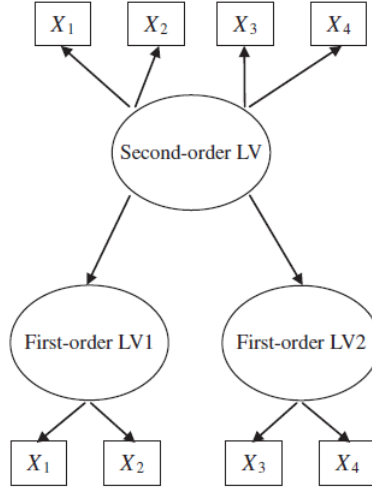


Fig. 4.2: Approccio ad indicatori ripetuti

L'approccio ad indicatori ripetuti può essere specificato considerando le seguenti tre equazioni:

$$\xi_{q,1}^I = \mathbf{B}_{q,q} * \xi_{q,1}^{II} + \zeta_{q,1} \quad (4.1)$$

$$\mathbf{x}_{p,1} = \Lambda_{p,q}^I * \xi_{q,1}^I + \delta_{p,1} \quad (4.2)$$

$$\mathbf{x}_{p,1} = \Lambda_{p,q}^{II} * \xi_{q,1}^{II} + \epsilon_{p,1} \quad (4.3)$$

dove gli indici m e p sono rispettivamente il numero di VL di primo ordine e di VM nel modello e l'indice q è il numero di VL di secondo ordine. I vettori ξ^I , ξ^{II} , \mathbf{x} , ζ , δ e ϵ indicano rispettivamente il primo e il secondo ordine di VL, le VM, gli errori strutturali e di misurazione. Le matrici \mathbf{B} , λ^I e λ^{II} definiscono i *path coefficient* che collegano le VL ed i fattori che legano rispettivamente le VM alle VL di primo ordine e di secondo ordine. Il modello strutturale o interno (3.1) specifica le relazioni tra il primo ordine e il secondo ordine. Le equazioni 3.2 e 3.3 indicano i modelli di misurazione, in cui le VM, misurando ciascuna VL di primo ordine, vengono ripetute per rappresentare il costrutto ad alto ordine.

Questo approccio è quello più utilizzato dai ricercatori quando si usa il PLS per la modellizzazione dei costrutti di ordine superiore, per la sua

semplicità e anche per il fatto che è stato presentato in modo più chiaro da prominenti metodologi del PLS (ad esempio Wold e Lohmöller).

Il vantaggio dell'approccio ad indicatori ripetuti è la sua capacità di stimare tutti i costrutti contemporaneamente, invece di stimare separatamente i costrutti di basso ordine e di alto ordine. Pertanto, prende in considerazione l'intera rete nomologica, non solo il livello più basso o il modello di livello superiore, evitando così confusione interpretativa. Quando si utilizza l'approccio ad indicatori ripetuti, i ricercatori devono prendere decisioni riguardo alla modalità di misurazione per il costrutto di alto ordine e lo schema di ponderazione interno. Alcuni autori elencano le linee guida per l'utilizzo di tipi diversi di modelli [8]. Innanzitutto, come per qualsiasi costrutto in un modello PLS-PM, il modo di misura per gli indicatori ripetuti ad alto ordine deve essere specificato (cioè modo A o modo B). Di solito, la misura del modo A è associata a costrutti riflessivi e il modo B è associato a costrutti formativi [73]; [187]. L'approccio standard per gli indicatori ripetuti sul modello di un costrutto di alto ordine è quello di utilizzare il modo A [208] che generalmente si adatta meglio ai modelli di tipo riflessivo-riflessivo. I tipi di modello formativo spesso sono anche stimati utilizzando il modo A per gli indicatori ripetuti, specialmente quando i costrutti del primo ordine sono riflessivi (cioè il tipo riflessivo-formativo) [25]; [83], anche se la natura formativa del costrutto di alto ordine potrebbe suggerire la misura in modo B. Pertanto, la maggior parte dei ricercatori pensa che sia più appropriato utilizzare il modo B per gli indicatori ripetuti di un modello di VL gerarchico di tipo formativo (cioè i tipi riflessivi-formativi e quelli formativi-formativi). Tuttavia, l'importanza della modalità di misurazione non è di solito discussa nei documenti di ricerca che presentano l'approccio ad indicatori ripetuti, ma solo indirettamente dedotti dalla direzione delle frecce nel *path diagram* [25]; [83]. In secondo luogo, oltre alla modalità di misurazione, Lohmöller [108] analizza analiticamente come impostare lo schema di ponderazione interna (ponderazione del fattore o del *path*) insieme alla modalità (modo A, modo B o entrambi) per i diversi tipi di VL gerarchici. Quindi, i ricercatori devono essere consapevoli che il tipo di schema di ponderazione interno che scelgono può dare una differenza importante nei risultati di un modello ad indicatori ripetuti; tuttavia Lohmöller [108] non fornisce alcuna indicazione su quale di queste impostazioni sia più adatta per ogni tipo. Inoltre, è spesso menzionato che l'approccio ad indicatori ripetuti è consigliato solo se i costrutti di basso ordine hanno un numero uguale di indicatori, in quanto altrimenti porterà a *loading*/pesi distorti per i costrutti di basso ordine sul costrutto di alto ordine [31]; [108]; [83]. Tuttavia, nella letteratura sembra mancare una valutazione

di questa assunzione generale.

Uno svantaggio di questo approccio è che esiste un effetto percepito di probabile distorsione delle stime utilizzando insieme variabili dello stesso tipo mediante la stima PLS. Secondo Rajala e Westerlund [139], l'approccio ad indicatori ripetuti può essere applicato a condizione che tutte le relazioni di misura siano di tipo riflessivo. Possono essere ipotizzate anche le relazioni strutturali formative del primo ordine su VL di secondo ordine, come è stato dimostrato in diversi studi [62]; [117]. Inoltre, l'uso ripetuto degli stessi indicatori può causare residui correlati artificialmente [8].

4.3.2 Approccio a due step

Un altro modo per costruire un modello di ordine superiore è quello di utilizzare l'approccio a due step: in questo approccio i punteggi delle VL vengono inizialmente stimati in un modello senza il costrutto del secondo ordine [107]. Una volta calcolati, i punteggi delle VL del primo ordine vengono successivamente utilizzati come indicatori della VL di alto ordine. Le VL del primo ordine sono quindi una combinazione lineare del costrutto di alto ordine, mentre le variabili osservate sono direttamente correlate solo alle dimensioni specifiche. Per questo viene definito approccio a due step.

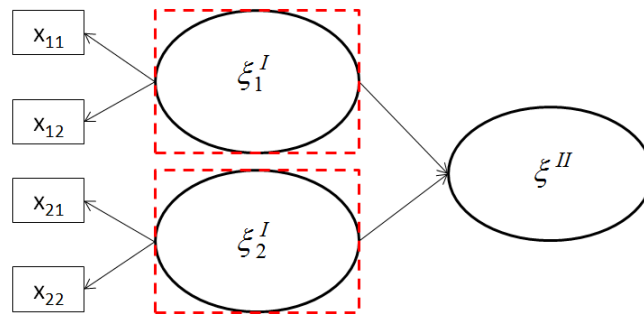


Fig. 4.3: Approccio a due step

Un tale approccio può offrire vantaggi nella stima dei modelli di alto ordine con indicatori formativi [43]; [142]. L'implementazione non viene eseguita tramite un'unica esecuzione PLS; ciò implica che qualsiasi costrutto di secondo ordine, esaminato nella fase 2, non viene preso in considerazione durante la stima dei punteggi delle VL nel primo step. Il primo passo di stima viene effettuato considerando solo il modello di misurazione che for-

nisce la stima dei costrutti di primo ordine, come riportato nella seguente equazione:

$$x_{p,1} = \Lambda_{p,q}^I * \xi_{q,1}^I + \delta_{p,1} \quad (4.4)$$

Nel secondo step, i punteggi stimati $\hat{\xi}^I$, ottenuti al primo step, vengono utilizzati come indicatori dei costrutti del secondo ordine:

$$\hat{\xi}_{q,1}^I = \mathbf{B}_{q,1} * \xi_{1,1}^{II} + \zeta_{q,1} \quad (4.5)$$

Sanchez [160] suggerisce questo metodo: possiamo ottenere un punteggio per un costrutto di primo ordine prendendo la prima componente principale dei suoi indicatori. Successivamente, i punteggi dei costrutti di basso ordine, così ottenuti, vengono utilizzati come indicatori per il costrutto di alto ordine in un modello di PLS-PM separato.

Quando ci si avvale dell'approccio a due fasi, si utilizza di solito la modalità di misura per il costrutto di ordine superiore nella seconda fase, che corrisponde all'operativizzazione del costrutto (cioè, modo B per un costrutto formativo e modo A per un costrutto riflessivo). L'approccio a due step ha il vantaggio di stimare un modello più parsimonioso sull'analisi del costrutto di alto ordine senza necessità di costrutti di basso ordine. Mentre, uno svantaggio chiaro di questo approccio a due step è che per qualsiasi costrutto che viene studiato, la fase due non viene presa in considerazione durante la stima dei punteggi delle VL nello stadio uno. Questo potrebbe causare un "interpretazione contraddittoria" [16]; [200]. Argomenti simili hanno seguito l'uso dell'approccio di modellizzazione a due fasi, come per esempio sostenuto da Anderson e Gerbing [4] nella letteratura CB-SEM. L'implementazione non è un'operazione PLS simultanea.

Un'altra differenza importante tra gli approcci emerge quando le VL gerarchiche vengono utilizzate in una rete nomologica di VL come un costrutto endogeno (cioè una conseguenza o un criterio). Quando si utilizza l'approccio ad indicatori ripetuti, indipendentemente dal tipo di misura, modo A o modo B, e il costrutto di alto ordine è formativo (cioè, formativo-riflessivo o formativo-formativo), i costrutti di basso ordine già spiegano tutta la varianza del costrutto di alto ordine (vale a dire, R^2 equivale a 1). Di conseguenza, altri costrutti antecedenti non possono spiegare alcuna varianza del costrutto di ordine superiore e i loro *path* del costrutto di alto ordine saranno zero (non significativi) [83]; [198]. Questo problema non si verifica quando viene utilizzato l'approccio a due step per i modelli forma-

tivi di alto ordine [83]; [8].

Alcuni studi si sono concentrati sul confronto tra i due approcci e sono limitati al caso di misure riflessive [112]; [200]. Da una prospettiva teorica, i due approcci portano a diverse definizioni del costrutto di secondo ordine. La differenza sta nel livello della distinzione tra la misurazione ed i modelli strutturali. Mentre negli approcci ad indicatori ripetuti le VL di ordine superiore sono direttamente misurate dall'intero insieme di VM (che a loro volta, misurano i fattori specifici del primo ordine), nel metodo a due step il costrutto di secondo ordine viene misurato direttamente con le VL del primo ordine.

Nel primo caso, il costrutto generale può essere visto come una variabile di contesto e il suo significato è indipendente dai rapporti con i fattori di primo ordine. Questa formalizzazione potrebbe essere applicata quando, ad esempio, si desidera valutare gli effetti di una modifica della percezione che era avvenuta in una VL del secondo ordine sulle VL di primo ordine o, nel caso di rapporti formativi, gli effetti di una percezione che cambia tra le VL del primo ordine sulle VL del secondo ordine. Pertanto, il modello ad indicatori ripetuti misurano l'intensità delle relazioni causali tra le sottomisure (VL di primo ordine) ed il contesto. Al contrario, con l'approccio a due step, il significato del costrutto di secondo ordine è definito dai rapporti con le sottomisure; cioè è misurato e non può esistere prima della stima delle VL del primo ordine. Le relazioni riflettono o formano la composizione della VL di alto ordine, anzi non rappresentano quanto le VL del primo ordine influenzino le VL del secondo ordine, ma la misura in cui i costrutti del primo ordine riflettono o formano il livello più alto di astrazione. Quindi, la differenza è nella direzione dell'impatto delle VL del secondo ordine sulle variabili osservate. Mentre l'approccio ad indicatori ripetuti collega direttamente le VL di secondo ordine sia alle VL del primo ordine che alle VM, nella stima a due step il costrutto generale ha effetti diretti sulle sotto dimensioni e solo degli effetti indiretti sulle VM. In un recente studio, Wilson et al. [199] hanno mostrato che l'affidabilità del costrutto di secondo ordine non dipende dall'approccio adottato. Ad ogni modo, l'approccio ad indicatori ripetuti produce stime polarizzate e meno coerenti (nel caso di piccoli campioni) rispetto all'approccio a due fasi.

In generale, la scelta del miglior approccio dipende chiaramente dal tipo di disegno. Nel caso in cui un costrutto di secondo ordine sia correlato in modo formativo alle dimensioni del primo ordine e ciascun costrutto sia misurato in modo riflessivo dalle sue VM, l'approccio a due step funziona meglio dell'altro approccio. Per quanto riguarda la quantità di varianza spiegata, l'approccio a due step produce relazioni meglio spiegate tra i due

ordini del modello. Inoltre, per quanto riguarda la stima dei parametri, in generale, l'approccio a due step è il migliore.

Possiamo concludere che per l'approccio ad indicatori ripetuti, la VL del secondo ordine, essendo gerarchicamente superiore, potrebbe essere vista come una variabile di contesto e l'attenzione si focalizza sull'impatto delle VL del primo ordine sulla VL di ordine superiore. Nell'approccio a due step, la VL del secondo ordine è misurata dalle VL del primo ordine e l'obiettivo è capire fino a che punto ogni VL del primo ordine riflette (in termini di covarianza) la composizione del livello del secondo ordine. Inoltre, l'approccio a due step risulta adatto per la stima dei costrutti formativi di secondo ordine poiché produce stime migliori rispetto a quelle ottenute con l'approccio ad indicatori ripetuti. L'approccio a due step è più coerente dal punto di vista teorico rispetto all'approccio ad indicatori ripetuti nella definizione del modello di misurazione della VL di secondo ordine. In conclusione, i costrutti misurati in modo riflessivo richiedono indicatori omogenei e l'approccio a due step, utilizzando una componente della VL anziché l'intero set di VM, riduce l'eterogeneità negli indicatori.

4.4 Nuovi metodi per costrutti di alto ordine nel PLS-PM

L'importanza della modellizzazione e della stima del costrutto di ordine superiore, sia dal punto di vista teorico che empirico, è stata riconosciuta da molti ricercatori sin dagli albori dell'analisi fattoriale [77]; [169] e recentemente sottolineato in molti studi [47]; [91]; [111]. Sfortunatamente, la ricerca è condotta quasi esclusivamente nell'area della SEM basata sulla covarianza. Tuttavia, lo scopo di stimare il costrutto di ordine superiore può essere raggiunto anche per mezzo del PLS-PM [107]; [208]. In letteratura esistono due diversi approcci che consentono di modellare e stimare i costrutti di alto ordine e la loro relazione con altri costrutti in una rete nomologica. Nei paragrafi precedenti questi approcci al costrutto di ordine superiore sono stati descritti in dettaglio e sono stati discussi alcuni dei loro limiti. Il primo approccio è stato proposto da Lohmöller [107], che ha suggerito di costruire il costrutto di ordine superiore come un costrutto generale collegato a tutte le VM del costrutto di ordine inferiore. Questo è il motivo per cui questo approccio è chiamato ad *indicatori ripetuti*. Il secondo approccio propone di stimare i punteggi del costrutto del primo ordine e poi usarli come indicatori per misurare il costrutto di ordine superiore: questo metodo viene quindi costruito in *due step*, con effetti diretti su ciascuno dei costrutti di primo ordine ed effetti indiretti sulle VM.

Tutti questi approcci possono essere applicati senza alcun problema computazionale ogni volta che le relazioni, ad ogni livello nel modello, sono riflesse. Ciò implica che le VM dovrebbero essere viste come causate dalle VL, come se riflettessero le VL del secondo ordine. I problemi di identificazione sorgono quando il ricercatore punta a misurare i costrutti formativi del secondo ordine, il che significa che sono concettualizzati come la combinazione delle loro sotto-dimensioni (le VL del primo ordine). Mentre il primo approccio consente la stima di questo tipo di modello, il secondo fallisce nell'identificazione di altre relazioni nel modello con il costrutto del secondo ordine. Come già accennato nei paragrafi precedenti questi approcci hanno dei limiti. In questi paragrafi ci soffermeremo dunque nel definire due nuovi approcci al fine proprio di superare alcuni di questi limiti; nello specifico verranno proposti due metodi alternativi per stimare i costrutti di ordine superiore, in particolare viene vedremo l'**approccio misto** per risolvere il problema relativo al potere predittivo della componente per ogni costrutto di ordine inferiore e l'**approccio con regressione PLS** per quanto riguarda la scelta del numero di componenti per ogni blocco.

4.4.1 *Approccio Misto*

Come già affermato, Sanchez [160] suggerisce l'approccio misto tramite il quale possiamo ottenere un punteggio per un costrutto di primo ordine prendendo la prima componente principale dei suoi indicatori. Successivamente, i punteggi dei costrutti di basso ordine, così ottenuti, vengono utilizzati come indicatori per il costrutto di alto ordine in un modello di PLS-PM separato. Questa componente cattura in modo ottimale la struttura della variabilità del blocco in modo da massimizzare la rappresentatività del blocco. Il suo limite è che in un *path modeling*, in cui vengono considerate tutte le relazioni tra le VL, non è in grado di prevedere la VL endogena. Per questo motivo, è stato proposto [20] un approccio alternativo per il calcolo delle VL di basso ordine.

Implementazione del metodo misto in PLS-PM

L'approccio misto inizia con l'implementazione del PLS-PM con un approccio ad indicatori ripetuti. In questo modo, l'algoritmo fornisce i punteggi dei blocchi di ordine inferiore. Quindi i punteggi dei blocchi vengono usati come indicatori del costrutto di alto ordine e, a questo punto, viene ripreso l'algoritmo PLS-PM. Possiamo schematizzare l'approccio misto scomponendolo in due fasi:

- In un primo momento viene costruito un costrutto di alto ordine da tutte le VM dei costrutti di ordine inferiore e viene eseguito l'algoritmo PLS-PM;
- I punteggi per ogni blocco, ottenuti dopo l'implementazione dell'algoritmo, vengono utilizzati come VM per il costrutto di alto ordine e l'algoritmo PLS-PM viene eseguito nuovamente.

Di seguito vengono descritti in dettaglio questi passaggi, considerando solo due livelli di costrutti. Poiché inizialmente il costrutto del secondo ordine non ha proprie VM, lo consideriamo formato da tutte le VM dei costrutti del primo ordine, come nella figura 4.4

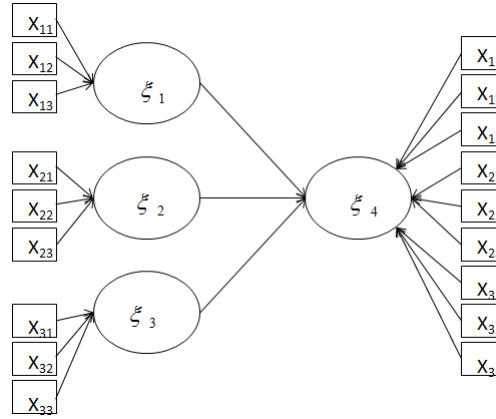


Fig. 4.4: Costrutto di secondo ordine con tutte le MV del primo ordine

In un primo momento, il modello esterno dei costrutti di primo ordine è espresso dall'equazione classica del PLS-PM:

$$\xi_q^I = \sum_{p=1}^{P_q} \omega_{pq} x_{pq} + \delta_q \quad (4.6)$$

mentre il modello strutturale, che specifica le relazioni tra le VL del primo ordine ed il costrutto del secondo ordine, è rappresentato dalla seguente equazione:

$$\xi_j^{II} = \sum_{(q: \xi_q^I \rightarrow \xi_j^{II})} \beta_{qj} \xi_q^I + \zeta_j \quad (4.7)$$

dove ξ_j^{II} è formato da tutte le VM del primo ordine:

$$\xi_j^{II} = \sum_{p=1}^{P_q} \omega_{pq} x_{pq} + \delta_q \quad (4.8)$$

Partendo da questa struttura, viene eseguito un algoritmo PLS-PM in modo da ottenere i punteggi di ogni blocco.

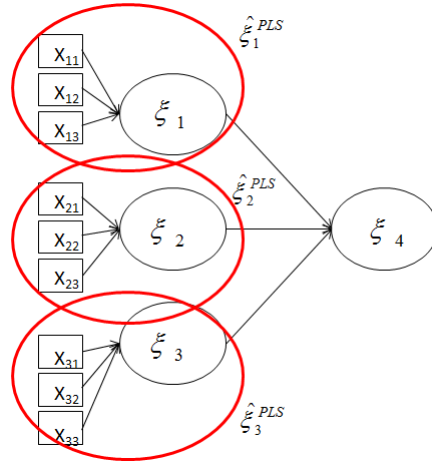


Fig. 4.5: Punteggi di ogni singolo costrutto di basso ordine

Una volta ottenuti i punteggi per ogni blocco, questi diventeranno le VM del costrutto di secondo ordine.

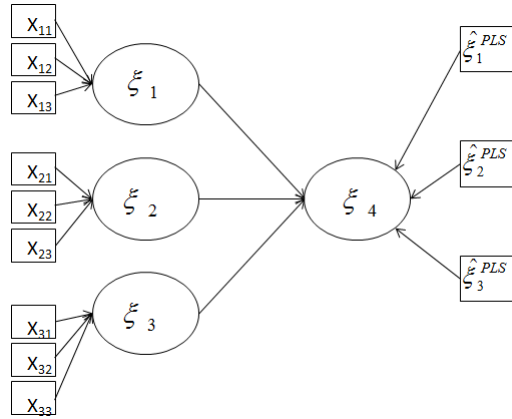


Fig. 4.6: Costrutto di secondo ordine con i punteggi del primo ordine

L'equazione del modello esterno del costrutto del primo ordine e l'equazione del modello strutturale sono le stesse di prima:

$$\xi_q^I = \sum_{p=1}^{P_q} \omega_{pq} x_{pq} + \delta_q \quad (4.9)$$

$$\xi_j^{II} = \sum_{(q: \xi_q^I \rightarrow \xi_j^{II})} \beta_{qj} \xi_q^I + \zeta_j \quad (4.10)$$

mentre l'equazione del modello esterno del costrutto di secondo ordine diventa una funzione dei componenti ottenuti:

$$\xi_j^{II} = \sum_{q=1}^Q \omega_h \hat{\xi}_q^I + \delta_j \quad (4.11)$$

A questo punto, una volta che i punteggi sono assegnati come indicatori del costrutto di secondo ordine, l'algoritmo PLS-PM può essere implementato.

Questo metodo viene proposto per utilizzare la componente che meglio rappresenta il suo blocco e, allo stesso tempo, ha un miglior potere predittivo sulla VL di alto ordine.

4.4.2 Approccio con regressione PLS

Come è già stato detto nei paragrafi precedenti, Sanchez [160], nel metodo a due step prende in considerazione solo la prima componente principale dei blocchi delle VL di basso ordine. Scegliendo solo la prima componente, può accadere che la parte rimanente della variabilità del blocco non sia presa in considerazione. Per questo motivo, viene proposto l' **approccio con regressione PLS** che può superare il problema relativo al numero di componenti dei costrutti di ordine inferiore e che consente di scegliere il numero di componenti da utilizzare (manualmente o secondo un criterio). Inoltre, poiché l'obiettivo del PLS-PM è quello di stimare le relazioni tra le VL, questo approccio fornisce componenti che rappresentano i loro blocchi e contemporaneamente sono predittivi del costrutto di ordine superiore.

il Metodo della regressione PLS

La regressione PLS è il metodo a cui la maggior parte degli studiosi pensa quando sente l'acronimo PLS. In breve, la regressione PLS è solo un algoritmo per l'analisi di regressione in cui vogliamo analizzare un blocco di

variabili di risposta Y in termini di un altro blocco di variabili predittive X . Quando abbiamo più di una variabile di risposta, si parla di PLS-R2, che rappresenta la versione PLS per la regressione multivariata.

Questa tecnica consente di associare un insieme di variabili predittive a una o più variabili di risposta. Contemporaneamente, il PLS-R decompone la matrice predittiva estraendo sequenzialmente componenti ortogonali che allo stesso tempo riassumono le variabili esplicative e permettono di modellare e predire le variabili di risposta. Il PLS-R può essere incluso tra i metodi di regressione regolarizzati, poiché gli stimatori PLS devono essere dimostrati come stimatori di contrazione [40].

Dal punto di vista algoritmico, la regressione PLS può essere considerata come un'estensione dell'algoritmo NIPALS all'analisi di una matrice di cross-covarianza. Inoltre, può essere considerata come una versione leggermente modificata dell'algoritmo PLS-PM a due blocchi.

Dato x_1, x_2, \dots, x_P un insieme di variabili predittive P e y_1, y_2, \dots, y_R , un insieme di variabili di risposta R misurate in N osservazioni e supponendo che tutte le variabili siano centrate, il modello PLS-R presuppone che vi sia una struttura comune sottostante i due blocchi di variabili e che questa struttura possa essere ripresa da poche componenti latenti t_h ($h = 1, \dots, H$), calcolate come una combinazione lineare delle variabili predittive. I predittori e le matrici di risposta \mathbf{X} e \mathbf{Y} vengono decomposti come:

$$\mathbf{X} = \mathbf{T}_H \mathbf{P}_H' + \mathbf{E}_H \quad (4.12)$$

$$\mathbf{Y} = \mathbf{T}_H \mathbf{C}_H' + \mathbf{F}_H \quad (4.13)$$

dove \mathbf{P}_H e \mathbf{C}_H sono le matrici dei *loadings*, \mathbf{E}_H e \mathbf{F}_H sono le matrici residue che rappresentano la parte della variabilità dei dati a causa del rumore. I parametri del modello sono calcolati mediante l'algoritmo di regressione PLS chiamato anche PLS-R2 nel caso di risposta multipla e PLS-R1 nel caso di risposta singola. Per una dettagliata rassegna delle proprietà matematiche e dell'algoritmo di PLS-R, si fa riferimento a Tenenhaus [184].

Dal punto di vista computazionale, l'algoritmo PLS-R può estrarre un numero di componenti uguale al rango di \mathbf{X} . Tuttavia, il modello di regressione PLS suppone che le informazioni comuni riportate da matrici di \mathbf{X} e \mathbf{Y} possano essere riassunte in poche componenti latenti. Quindi, un problema cruciale nel modello PLS-R è la definizione del numero H di componenti da conservare. Nella regressione PLS la capacità esplicativa del modello (misurata in termini di indice R^2) aumenta finché il numero delle componenti aumenta. Al contrario, l'abilità predittiva del modello, intesa come capacità esplicativa del modello, riferita a unità che non sono state considerate

nella costruzione del modello (set di convalida), comincia a diminuire dopo un certo numero di componenti. Ciò significa che il modello sovrappone i dati ed è necessario smettere di estrarre le componenti. Di solito viene eseguita una procedura di convalida incrociata per valutare se la componente h -esima aumenta la capacità predittiva del modello. Il campione originale viene suddiviso in S sottocampioni. Per S , un sottoinsieme diverso viene mantenuto come dati di convalida e i restanti $(S - 1)$ sottocampioni vengono utilizzati come dati formativi. Ogni volta, per ogni unità del set di convalida, vengono calcolati gli errori quadratici di previsione $e_{(-i)r}^2$ riferiti a \mathbf{y}_r . Per ogni modello h - componente, l'indice di predizione della somma dei quadrati degli errori (PRESS) viene ottenuto come:

$$PRESS_{rh} = \sum e_{(-i)r}^2 \quad (4.14)$$

Il sovradimensionamento del modello viene studiato graficando l'indice PRESS contro il numero delle componenti. In genere, il PRESS diminuisce per un certo numero di componenti, poi comincia ad aumentare: ovviamente, si sceglie il numero di componenti che danno il minimo PRESS. Per misurare il contributo marginale della componente h -esima alla potenza predittiva del modello si utilizza l'indice Q^2 [5]

$$Q_h^2 = 1 - \frac{\sum_{r=1}^R PRESS_{rh}}{\sum_{r=1}^R RESS_{r(h-1)}} \quad (4.15)$$

dove $RESS_{hr}$ è la somma dei residui quadrati di \mathbf{y}_r in un modello di componenti $h - 1$ nell'insieme di dati. Non ci sono test *ad hoc* per valutare il significato di questo indice; nella pratica, la componente h -esima viene mantenuta se $Q_h^2 \geq 0,975$.

L'equazione della regressione

La regressione PLS fornisce un'equazione di regressione classica, in cui la risposta è stimata come una combinazione lineare delle variabili predittive. La seguente equazione può essere derivata dall'ultimo passaggio dell'algoritmo PLS-R:

$$\mathbf{Y} = \mathbf{t}_1 \mathbf{c}'_1 + \mathbf{t}_2 \mathbf{c}'_2 + \mathbf{t}_H \mathbf{c}'_H + \mathbf{F}_H = \mathbf{T}_H \mathbf{C}'_H \quad (4.16)$$

Questa è l'equazione di regressione di un modello PLS-R di una componente H , in cui le variabili di risposta sono espresse in funzione delle componenti PLS. Nell'algoritmo PLS-R ogni \mathbf{t}_h viene calcolato in funzione di \mathbf{E}_{h-1}

$$t_h = E_{h-1}\omega_h \quad (4.17)$$

In un modello con H componenti, la matrice \mathbf{T}_H dei fattori degli \mathbf{X} – *punteggi* possono essere ottenuti in funzione delle variabili originali \mathbf{X} . Dopo alcune sostituzioni, otteniamo le risposte come una funzione lineare delle variabili predittive:

$$\mathbf{Y} = \mathbf{T}_H \mathbf{C}'_H + \mathbf{F}_H = \mathbf{X} \mathbf{B}_H^{PLS} + \mathbf{F}_H \quad (4.18)$$

dove \mathbf{B}_H^{PLS} è la matrice dei coefficienti di un modello di regressione PLS di H componenti.

L'implementazione della regressione PLS per il PLS-PM di alto ordine

Il modello PLS-R presuppone che esista una struttura comune sottostante i due blocchi di variabili e che tale struttura possa essere ripresa da poche componenti latenti $t_h (h = 1...H)$, calcolata come una combinazione lineare delle variabili predittive.

Nel caso del costrutto di alto ordine, i costrutti di ordine inferiore sono considerati come blocchi di variabili predittive ed il costrutto di alto ordine come il blocco di variabili di risposta. In questo modo viene eseguita una regressione PLS per ogni blocco, in modo da ottenere H componenti per ogni blocco. Quindi queste H componenti rappresenteranno le VM del costrutto di alto ordine.

L'approccio con regressione PLS può essere schematizzato in tre fasi:

- In un primo momento, un costrutto di ordine superiore è formato da tutte le VM dei costrutti di ordine inferiore;
- la regressione PLS viene applicata per ottenere H componenti per ogni blocco;
- Una volta ottenute le H componenti, queste saranno utilizzate come VM del costrutto di alto ordine su cui viene eseguito l'algoritmo PLS-PM.

Questo metodo, al momento, è applicato solo per i costrutti di secondo livello. Di seguito verranno descritti dettagliatamente i passaggi, considerando solo due livelli di costrutti. Quindi, inizialmente, poiché il costrutto del secondo ordine non ha le proprie VM, lo consideriamo formato da tutte le VM dei costrutti del primo ordine, come nella figura 4.7

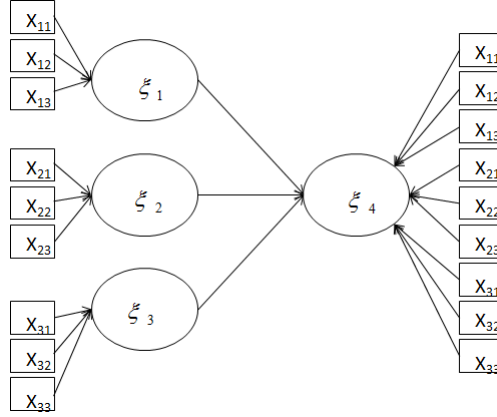


Fig. 4.7: Costrutto di secondo ordine con tutte le MV del primo ordine

In un primo momento, il modello esterno dei costrutti di primo ordine è espresso dall'equazione classica del PLS-PM:

$$\xi_q^I = \sum_{p=1}^{P_q} \omega_{pq} x_{pq} + \delta_q \quad (4.19)$$

mentre il modello strutturale, che specifica le relazioni tra le VL del costrutto del primo ordine e del costrutto di secondo ordine, è rappresentato dalla seguente equazione:

$$\xi_j^{II} = \sum_{(q: \xi_q^I \rightarrow \xi_j^{II})} \beta_{qj} \xi_q^I + \zeta_j \quad (4.20)$$

dove ξ_j^{II} è formato da tutte le VM del primo ordine:

$$\xi_j^{II} = \sum_{p=1}^{P_q} \omega_{pq} x_{pq} + \delta_q \quad (4.21)$$

A partire da questa struttura, la regressione PLS viene applicata per ogni blocco del costrutto di primo ordine, in cui ogni blocco del primo ordine rappresenta un insieme di variabili predittive e il costrutto di secondo ordine è un insieme di variabili di risposta.

Una volta ottenute le H componenti per i blocchi, queste saranno le VM del costrutto di secondo ordine:

L'equazione del modello esterno del costrutto del primo ordine e l'equazione del modello strutturale sono le stesse:

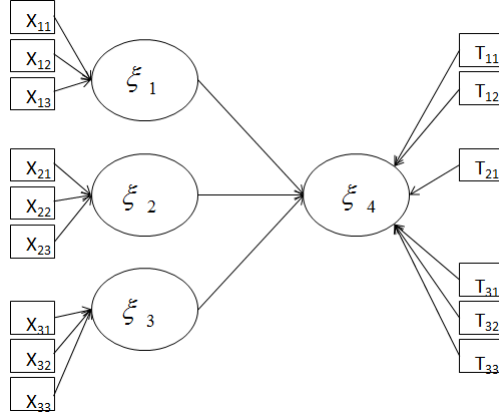


Fig. 4.8: Costrutto di secondo ordine con le componenti PLS-R del primo ordine

$$\xi_q^I = \sum_{p=1}^{P_q} \omega_{pq} x_{pq} + \delta_q \quad (4.22)$$

$$\xi_j^{II} = \sum_{(q: \xi_q^I \rightarrow \xi_j^{II})} \beta_{qj} \xi_q^I + \zeta_j \quad (4.23)$$

mentre l'equazione del modello esterno del costrutto di secondo ordine diventa una funzione delle componenti ottenute:

$$\xi_j^{II} = \sum_{h=1}^H \omega_h T_h^I + \delta_j \quad (4.24)$$

A questo punto, una volta che le componenti PLS-R vengono assegnate come indicatori del costrutto del secondo ordine, l'algoritmo PLS-PM può essere implementato.

Questo approccio è proposto per superare l'inconveniente dell'approccio a due step relativo al numero di componenti scelte nei costrutti di primo ordine, al fine di acquisire la maggior variabilità del blocco possibile.

5. STUDIO DI SIMULAZIONE

Lo scopo di questo studio è di indagare, nello stesso disegno di simulazione, le prestazioni di un modello PLS-PM di alto ordine attraverso l'utilizzo di tre modelli di stima differenti: un approccio a variabili ripetute, un approccio misto ed un approccio con regressione PLS-PM. Inoltre i tre approcci sono stati utilizzati e comparati attraverso le simulazioni in differenti situazioni modellistiche:

- Modello di misurazione formativo e modello strutturale formativo
- Modello di misurazione formativo e modello strutturale riflessivo
- Modello di misurazione riflessivo e modello strutturale formativo
- Modello di misurazione riflessivo e modello strutturale riflessivo

L'obiettivo della simulazione è quello di confrontare le prestazioni dei vari approcci sulla base di campioni di numerosità differenti, al fine di comprendere l'effetto della dimensione degli stesso. Le prestazioni sono valutate attraverso l'utilizzo di varie misure quali: la precisione di previsione, la stima degli errori e l'efficienza degli approcci considerati. I paragrafi seguenti riportano il piano di simulazione ed i commenti sui risultati ottenuti.

5.1 *Generazione dati*

La simulazione Montecarlo è stata condotta con il pacchetto *boot* di R. Il processo di generazione dei dati è coerente con la procedura descritta da Paxton et al. per uno studio Montecarlo su modelli SEM [134]. In primo luogo si definisce la struttura del modello e dei parametri della popolazione; si procede poi generando casualmente la VL di secondo ordine e, dati i parametri ed i termini di errore, stimando le VL di primo ordine. Secondo i parametri esterni e i termini di errore, nell'ultimo passaggio, si generano le

VM di primo e secondo ordine. Il modello di popolazione sottostante utilizzato per la simulazione consiste in una VL di secondo ordine (indicata con ξ^{II}) ed n VL di primo ordine (indicate da $\xi_1^I, \xi_2^I, \xi_3^I, \dots, \xi_n^I$).

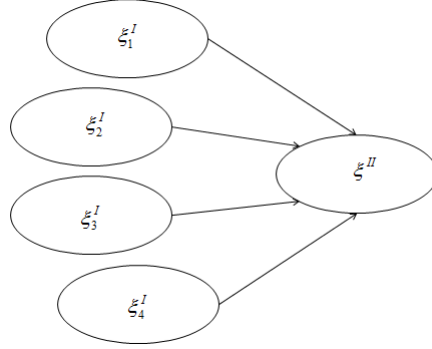


Fig. 5.1: Path Diagram

Le performance dei tre approcci sono state comparate sulla base di differenti numerosità campionarie ($n= 100, 250, 500, 1000$). Per ogni condizione sono state considerate 500 replicazioni.

Il costrutto di secondo ordine, in termini di numero di variabili, differisce a seconda dell'approccio di stima usato: per l'approccio ripetuto esso è formato da tutte le variabili del primo ordine, nell'approccio misto è formato da un numero di variabili uguale al numero di VL del primo ordine, infine per il modello con regressione PLS-PM la numerosità del blocco varia a seconda del numero di componenti considerate significative dall'algoritmo.

Per iniziare sono state generate le VL di primo ordine ξ_i^I come variabili casuali $\xi_q^I \sim N(0, 1)$ [36]. I dati generati sono stati riscaliati in un intervallo $[1, 100]$

Per il modello strutturale formativo il costrutto di secondo ordine ξ_j^{II} è stato calcolato come il prodotto di ξ_q^I per il vettore dei coefficienti del *path* β_{qj} con l'aggiunta di una componente di errore ζ_j secondo la seguente equazione:

$$\xi_j^{II} = \sum_{(q: \xi_q^I \rightarrow \xi_j^{II})} \beta_{qj} \xi_q^I + \zeta_j \quad (5.1)$$

Per il modello strutturale riflessivo i costrutti di primo ordine ξ_q^I sono stati calcolati come il prodotto di ξ_j^{II} per il vettore dei coefficienti del *pa-*

th β_{qj} con l'aggiunta di una componente di errore ζ_q secondo la seguente equazione:

$$\xi_q^I = \sum_{(q:\xi_j^{II} \rightarrow \xi_q^I)} \beta_{qj} \xi_j^{II} + \zeta_q \quad (5.2)$$

Si assume che il vettore dei coefficienti (β) del modello strutturale abbia elementi tutti uguali a 0,7.

Ogni vettore della componente di errore ζ viene estratto da una distribuzione normale univariata [78] con media uguale a zero e varianza, $var(\zeta)$, scelta per soddisfare il costrutto di secondo ordine, secondo l'equazione:

$$R_j^2 = \frac{var(modello_j)}{var(totale_j)} = \frac{var(modello_j)}{var(modello_j) + var(errore_j)} \quad (5.3)$$

dove la $var(totale_j)$ è la varianza di ξ_j^{II} , data da:

$$\xi_j^{II} = \xi_q^I \beta_{qj} + \zeta_j = modello_j + errore_j \quad (5.4)$$

e $var(\zeta_j)$ è:

$$var(\zeta_j) = \frac{var(\xi_q^I) * (1 - R_q^2)}{R_q^2} \quad (5.5)$$

Il valore dell' R^2 per il costrutto di secondo ordine è impostato a 0,7. Le VM sono generate a partire dalle VL, dati i coefficienti λ , con la seguente formula:

$$X_{(nq)} = \xi_q^I * (\lambda^I)_k^{-1} + \delta_{nq} \quad (5.6)$$

dove l'errore è distribuito come una variabile uniforme continua $\delta \sim (-1, 1)$.

Esistono poi tre misure comunemente riportate in letteratura che vengono utilizzate per valutare quanto bene i metodi stimano i parametri: il bias relativo (RB), la deviazione standard (StD) e l'errore quadratico medio ($RMSE$) delle stime.

RB è calcolato come:

$$RB = \frac{1}{n} \sum_{i=1}^n \frac{(\hat{\theta}_i - \theta)}{\theta} \quad i = 1, 2, \dots, 500 \quad (5.7)$$

dove n rappresenta il numero di repliche nella simulazione, $\hat{\theta}_i$ è la stima dei parametri per ogni replica e θ è il parametro della popolazione corrispondente. La formula è equivalente alla media *RB* [141]. Un *RB* positivo indica una sovrastima del vero parametro, un *RB* negativo una sottostima.

StD è calcolato come:

$$StD = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{\theta}_i - E(\hat{\theta}))^2} \quad i = 1, 2, \dots, 500 \quad (5.8)$$

dove $E(\hat{\theta})$ è la media delle stime dei 500 set di dati simulati; questo indice fornisce informazioni sull'efficienza delle stime.

Infine, l'*RMSE* è calcolato come:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{\theta}_i - \theta)^2} \quad i = 1, 2, \dots, 500 \quad (5.9)$$

$MSE = bias(\hat{\theta})^2 + Var(\hat{\theta})$. Pertanto, l'*RMSE* contiene informazioni sia sulla bias che sulla variabilità delle stime.

Nei paragrafi successivi si analizzeranno i risultati delle simulazioni effettuate con diversi schemi di variabili. Precisamente sono stati simulati tre modelli con:

- 3 VL di primo ordine ed una di secondo ordine con 8 VM per blocco.
- 4 VL di primo ordine ed una di secondo ordine con 5 VM per blocco.
- 7 VL di primo ordine ed una di secondo ordine con 4 VM per blocco.

Per ogni schema si è deciso di confrontare tra loro l'indice di comunaltà, in quanto è l'unico indice sempre calcolabile, con ogni tipo di relazione sia per il modello strutturale che per il modello di misura. Inoltre la comunaltà, per i modelli riflessivi è pari all'AVE al netto di una costante, invece per i modelli formativi si approssima all'indice di ridondanza. Si rimanda ai paragrafi 3.4.2 e 3.4.3 per i dettagli.

5.2 Risultati: Modello a 3 variabili latenti di primo ordine con 8 manifeste per blocco

La prima simulazione è stata effettuata considerando un modello con 3 VL di primo ordine composte ogniuna da 8 VM.

5.2.1 Confronto comunalità

La tabella 5.1 riporta i risultati relativi alla comunalità, alle bias e gli errori standard raggruppati in base all'approccio, alle relazioni del modello esterno ed interno e alla numerosità del campione.

Tab. 5.1: Confronto Comunalità

Approccio	Numerosità	Valore	Rif.-Form.	Form.-Rif.	Rif.-Rif.	Form.-Form.
Ripetuto	100	Comunalità	0,39605	0,36746	0,35992	0,41777
		Bias	0,00450	-0,00274	-0,00263	-0,00619
		Errore St.	0,03273	0,04836	0,03522	0,03019
	250	Comunalità	0,42624	0,42435	0,37188	0,34235
		Bias	0,00179	-0,00675	0,00105	-0,01025
		Errore St.	0,02093	0,02797	0,02158	0,03006
	500	Comunalità	0,39763	0,37871	0,40283	0,37789
		Bias	0,00142	-0,00638	0,00114	-0,00176
		Errore St.	0,01495	0,02034	0,01346	0,01637
	1000	Comunalità	0,37522	0,40788	0,41753	0,34613
		Bias	0,00039	-0,00155	0,00020	-0,00177
		Errore St.	0,01042	0,01066	0,01002	0,01316
Misto	100	Comunalità	0,72124	0,83639	0,67927	0,82546
		Bias	0,00628	0,02496	0,00072	0,02519
		Errore St.	0,03422	0,00848	0,03848	0,02135
	250	Comunalità	0,74293	0,80283	0,69336	0,75840
		Bias	0,00214	0,01148	0,00095	0,01199
		Errore St.	0,01978	0,01748	0,02294	0,01801
	500	Comunalità	0,71875	0,75757	0,72329	0,77743
		Bias	0,00222	0,00654	0,00236	0,00598
		Errore St.	0,01477	0,01364	0,01352	0,01340
	1000	Comunalità	0,69207	0,76046	0,73592	0,73452
		Bias	0,00082	0,00417	0,00035	0,00335
		Errore St.	0,01136	0,00984	0,00995	0,01004
PLS-R	100	Comunalità	0,97533	0,98933	0,96912	0,98891
		Bias	-0,06186	-0,00717	-0,01849	-0,01567
		Errore St.	0,08976	0,00963	0,04993	0,04809
	250	Comunalità	0,98828	0,98567	0,97379	0,97757
		Bias	-0,00109	-0,02477	-0,00299	-0,00393
		Errore St.	0,00417	0,06098	0,01256	0,00844
	500	Comunalità	0,98136	0,98286	0,98476	0,97703
		Bias	-0,00148	-0,00166	-0,00047	-0,00244
		Errore St.	0,01295	0,00436	0,00321	0,01556
	1000	Comunalità	0,97543	0,98672	0,98690	0,97413
		Bias	-0,00045	-0,00028	-0,00027	-0,00067
		Errore St.	0,00391	0,00204	0,00200	0,00423

La comunalità stimata è sempre significativa. Considerando il valore dell'indicatore, come si può notare in figura 5.2, esso è sempre più alto per l'algoritmo del PLS-R, per ogni numerosità del campione e per ogni tipologia di rapporto del modello. In tabella 5.1 si nota inoltre come per

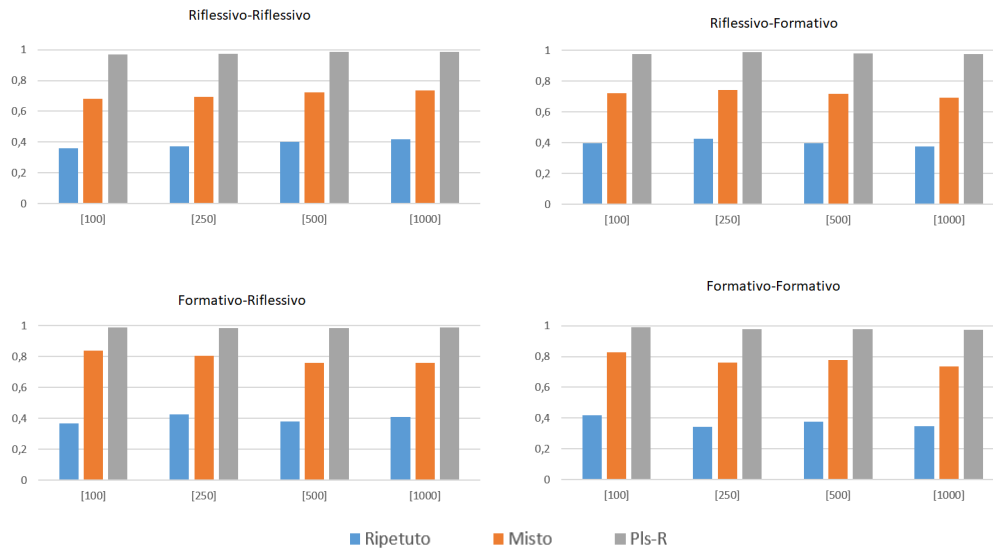


Fig. 5.2: Confronto Comunalità

l'approccio PLS-R, la variabilità della stima si riduca all'aumentare della dimensione del campione.

5.2.2 Path coefficients

Le seguenti tabelle riportano i risultati relativi ai coefficienti β , alle bias e gli errori standard raggruppati in base all'approccio, alle relazioni del modello esterno ed interno e alla numerosità del campione.

Come si può notare nelle tabelle 5.2, 5.3, 5.4, 5.5, al variare della numerosità campionaria e dell'approccio utilizzato, i valori dei diversi path non si discostano molto tra loro. Le stime dei path sono sempre significative ma, pur rimanendo bassa, la variabilità delle stime per l'approccio PLS-R è più alta rispetto a quella degli altri approcci

Tab. 5.2: Modello Formativo-Formativo

Approccio	Numerosità	Valore	β_1	β_2	β_3
Ripetuto	100	Path	0,36676	0,36681	0,36678
		Bias	-0,00527	-0,00530	-0,00529
		Errore St.	0,00451	0,00452	0,00452
	250	Path	0,38261	0,38257	0,38263
		Bias	-0,00274	-0,00273	-0,00274
		Errore St.	0,00451	0,00451	0,00452
	500	Path	0,37814	0,37815	0,37813
		Bias	-0,00148	-0,00149	-0,00149
		Errore St.	0,00323	0,00323	0,00323
	1000	Path	0,38903	0,38896	0,38890
		Bias	-0,00093	-0,00091	-0,00089
		Errore St.	0,00265	0,00265	0,00265
Misto	100	Path	0,36672	0,36681	0,36680
		Bias	-0,00525	-0,00531	-0,00528
		Errore St.	0,00452	0,00450	0,00453
	250	Path	0,38259	0,38254	0,38262
		Bias	-0,00275	-0,00270	-0,00274
		Errore St.	0,00451	0,00452	0,00450
	500	Path	0,37811	0,37810	0,37815
		Bias	-0,00146	-0,00149	-0,00151
		Errore St.	0,00324	0,00323	0,00323
	1000	Path	0,38906	0,38891	0,38886
		Bias	-0,00100	-0,00086	-0,00087
		Errore St.	0,00265	0,00265	0,00266
PLS-R	100	Path	0,32874	0,38836	0,38251
		Bias	-0,00148	-0,00768	-0,00616
		Errore St.	0,02656	0,02851	0,03202
	250	Path	0,40412	0,37802	0,36897
		Bias	-0,00378	0,00239	-0,00605
		Errore St.	0,01911	0,01639	0,01779
	500	Path	0,37983	0,34919	0,40257
		Bias	-0,00192	-0,00095	-0,00132
		Errore St.	0,01525	0,01501	0,00978
	1000	Path	0,40105	0,36301	0,40649
		Bias	-0,00036	-0,00112	-0,00112
		Errore St.	0,00697	0,00917	0,00767

Tab. 5.3: Modello Riflessivo-Riflessivo

Approccio	Numerosità	Valore	β_1	β_2	β_3
Ripetuto	100	Path	0,85078	0,81608	0,80438
		Bias	0,00047	-0,00551	0,00262
		Errore St.	0,02490	0,04015	0,04012
	250	Path	0,85933	0,78771	0,84767
		Bias	0,00081	0,00085	0,00038
		Errore St.	0,01691	0,02508	0,01826
	500	Path	0,83213	0,87268	0,84609
		Bias	0,00255	0,00018	0,00091
		Errore St.	0,01387	0,01095	0,01175
	1000	Path	0,84362	0,86307	0,86624
		Bias	0,00021	-0,00077	0,00063
		Errore St.	0,00901	0,00826	0,00791
Misto	100	Path	0,83163	0,82963	0,81146
		Bias	0,00087	-0,00448	0,00231
		Errore St.	0,02869	0,03489	0,03242
	250	Path	0,84825	0,79979	0,84778
		Bias	0,00058	0,00116	0,00055
		Errore St.	0,01619	0,02371	0,01634
	500	Path	0,84061	0,86039	0,85074
		Bias	0,00178	0,00136	0,00049
		Errore St.	0,01214	0,01080	0,01021
	1000	Path	0,85280	0,85974	0,86062
		Bias	0,00999	0,00016	0,00061
		Errore St.	0,00836	0,00798	0,00760
PLS-R	100	Path	0,84827	0,81675	0,80644
		Bias	0,00042	-0,00415	0,00111
		Errore St.	0,02708	0,03773	0,04022
	250	Path	0,85782	0,79082	0,84620
		Bias	0,00027	0,00122	0,00067
		Errore St.	0,01718	0,02320	0,01805
	500	Path	0,83303	0,87161	0,84628
		Bias	0,00241	0,00034	0,00090
		Errore St.	0,01331	0,01093	0,01149
	1000	Path	0,84416	0,86278	0,86591
		Bias	0,00028	0,00379	0,00052
		Errore St.	0,00874	0,00816	0,00781

Tab. 5.4: Modello Formativo-Riflessivo

Approccio	Numerosità	Valore	β_1	β_2	β_3
Ripetuto	100	Path	0,92425	0,91622	0,90345
		Bias	0,01115	0,01716	0,01147
		Errore St.	0,01199	0,00790	0,01051
	250	Path	0,90479	0,89695	0,88653
		Bias	0,00413	0,00724	0,00761
		Errore St.	0,01084	0,01130	0,01247
	500	Path	0,86688	0,87877	0,86621
		Bias	0,00377	0,00369	0,00302
		Errore St.	0,01022	0,00922	0,01082
	1000	Path	0,87446	0,87201	0,87014
		Bias	0,00235	0,00221	0,00222
		Errore St.	0,00714	0,00720	0,00695
Misto	100	Path	0,92421	0,91621	0,90343
		Bias	0,01117	0,01712	0,01151
		Errore St.	0,01194	0,00789	0,01053
	250	Path	0,90465	0,89690	0,88656
		Bias	0,00422	0,00725	0,00754
		Errore St.	0,01084	0,01130	0,01247
	500	Path	0,86684	0,87873	0,86612
		Bias	0,00375	0,00370	0,00304
		Errore St.	0,01021	0,00922	0,01082
	1000	Path	0,87439	0,87197	0,87005
		Bias	0,00236	0,00219	0,00224
		Errore St.	0,00714	0,00720	0,00695
PLS-R	100	Path	0,89227	0,89036	0,89544
		Bias	0,01931	0,01046	0,00802
		Errore St.	0,02116	0,01914	0,01187
	250	Path	0,89953	0,87616	0,88671
		Bias	0,00361	0,00431	0,00461
		Errore St.	0,01256	0,01504	0,01327
	500	Path	0,85724	0,86620	0,85966
		Bias	0,00231	0,00161	0,00051
		Errore St.	0,01161	0,01165	0,01133
	1000	Path	0,86234	0,86526	0,87088
		Bias	0,00137	0,00144	0,00158
		Errore St.	0,00747	0,00752	0,00709

Tab. 5.5: Modello Riflessivo-Formativo

Approccio	Numerosità	Valore	β_1	β_2	β_3
Ripetuto	100	Path	0,41027	0,40284	0,36349
		Bias	-0,00293	-0,00325	0,00166
		Errore St.	0,02362	0,01969	0,02195
	250	Path	0,42130	0,34822	0,38961
		Bias	-0,00066	-0,00091	-0,00016
		Errore St.	0,01312	0,01254	0,01158
	500	Path	0,42538	0,39293	0,35942
		Bias	-0,00113	-0,00032	-0,00023
		Errore St.	0,00972	0,01092	0,00926
	1000	Path	0,43767	0,37558	0,38726
		Bias	-0,00117	0,00044	0,00045
		Errore St.	0,00762	0,00732	0,00807
Misto	100	Path	0,39320	0,40693	0,37709
		Bias	-0,00258	0,00053	-0,00257
		Errore St.	0,01133	0,01338	0,00960
	250	Path	0,38944	0,38315	0,38763
		Bias	-0,00076	-0,00116	0,00035
		Errore St.	0,00685	0,00682	0,00661
	500	Path	0,40288	0,39053	0,38565
		Bias	-0,00111	0,00020	-0,00071
		Errore St.	0,00581	0,00505	0,00477
	1000	Path	0,40416	0,39742	0,39992
		Bias	-0,00012	-0,00032	0,00016
		Errore St.	0,00439	0,00401	0,00450
PLS-R	100	Path	0,40855	0,39913	0,36940
		Bias	-0,00678	-0,01608	0,01699
		Errore St.	0,05362	0,05742	0,05601
	250	Path	0,42115	0,34952	0,38845
		Bias	-0,00263	0,00053	0,00051
		Errore St.	0,01235	0,01167	0,01100
	500	Path	0,42256	0,39329	0,36202
		Bias	-0,00061	-0,00175	0,00064
		Errore St.	0,01098	0,01619	0,00995
	1000	Path	0,43400	0,37862	0,38799
		Bias	-0,00064	0,00018	0,00020
		Errore St.	0,00696	0,00672	0,00732

5.2.3 Relative Bias

Le seguenti tabelle riportano i risultati dei *relative bias* raggruppati in base all'approccio, alle relazioni del modello esterno ed interno e alla numerosità del campione.

Nella Tabella 5.6 troviamo i valori degli RB per i path del modello Formativo-Formativo.

Tab. 5.6: Relative Bias Modello Formativo-Formativo					
Approccio	Path	100	250	500	1000
Ripetuto	β_1	-0,01437	-0,00716	-0,00392	-0,00239
	β_2	-0,01445	-0,00714	-0,00395	-0,00233
	β_3	-0,01442	-0,00716	-0,00394	-0,00230
Misto	β_1	-0,01433	-0,00720	-0,00387	-0,00256
	β_2	-0,01449	-0,00705	-0,00393	-0,00222
	β_3	-0,01440	-0,00717	-0,00399	-0,00223
PLS-R	β_1	-0,00449	-0,00936	-0,00506	-0,00091
	β_2	-0,01978	0,00634	-0,00272	-0,00309
	β_3	-0,01611	-0,01639	-0,00328	-0,00275

Osservando gli RB si nota che essi sono sempre molto vicini allo zero, dunque in tutti gli approcci i *path* risultano ben stimati, per l'approccio PLS-R però la stima è leggermente più lontana dallo zero rispetto agli altri approcci, tranne che nel caso del campione da 1000 osservazioni, in cui l'RB di tutti i *path* è più vicino allo zero. Nella Tabella 5.7 troviamo i valori degli RB per i path del modello Riflessivo-Riflessivo.

Tab. 5.7: Relative Bias Modello Riflessivo-Riflessivo					
Approccio	Path	100	250	500	1000
Ripetuto	β_1	0,00056	0,00094	0,00307	0,00025
	β_2	-0,00675	0,00109	0,00021	-0,00001
	β_3	0,00327	0,00046	0,00108	0,00073
Misto	β_1	0,00105	0,00069	0,00212	0,00012
	β_2	-0,00540	0,00146	0,00159	0,00019
	β_3	0,00284	0,00065	0,00058	0,00072
PLS-R	β_1	0,00050	0,00033	0,00290	0,00033
	β_2	-0,00508	0,00155	0,00039	0,00004
	β_3	0,00139	0,00080	0,00106	0,00060

Tab. 5.8: Relative Bias Modello Riflessivo-Formativo

Approccio	Path	100	250	500	1000
Ripetuto	β_1	-0,00714	-0,00157	-0,00265	-0,00267
	β_2	-0,00807	-0,00262	-0,00081	0,00118
	β_3	0,00458	-0,00041	-0,00065	0,00117
Misto	β_1	-0,00657	-0,00195	-0,00275	-0,00030
	β_2	0,00131	-0,00301	0,00051	-0,00082
	β_3	-0,00683	0,00090	-0,00184	0,00041
PLS-R	β_1	-0,01661	-0,00625	-0,00145	-0,00148
	β_2	-0,04029	0,00151	-0,00446	0,00048
	β_3	0,04598	0,00131	0,00178	0,00053

Tab. 5.9: Relative Bias Modello Formativo-Riflessivo

Approccio	Path	100	250	500	1000
Ripetuto	β_1	0,01206	0,00456	0,00436	0,00269
	β_2	0,01874	0,00807	0,00420	0,00254
	β_3	0,01270	0,00859	0,00349	0,00255
Misto	β_1	0,01209	0,00467	0,00434	0,00271
	β_2	0,01869	0,00809	0,00422	0,00252
	β_3	0,01274	0,00851	0,00352	0,00258
PLS-R	β_1	0,02165	0,00402	0,00270	0,00160
	β_2	0,01175	0,00493	0,00186	0,00167
	β_3	0,00896	0,00519	0,00059	0,00182

Anche nel caso di una relazione Riflessiva-Riflessiva si nota che gli RB sono sempre molto vicino allo zero, per l'approccio PLS-R la stima risulta più adeguata per tutte le numerosità, tranne che nel caso del campione da 1000 osservazioni, in cui i *path* risultano leggermente sovrastimati rispetto agli altri approcci. Nella Tabella 5.8 troviamo i valori degli RB per i path del modello Riflessivo-Formativo.

In questo caso per l'approccio PLS-R il primo *path* viene sottostimato rispetto agli altri approcci, mentre all'aumentare della numerosità gli altri *path* vengono stimati meglio nel PLS-R. Nella Tabella 5.9 troviamo i valori degli RB per i path del modello Formativo-Riflessivo.

Anche in questo caso, con l'approccio PLS-R, il primo *path* viene sovrastimato rispetto agli altri approcci, mentre all'aumentare della numerosità gli altri *path* vengono stimati meglio con il PLS-R.

5.3 Risultati: Modello a 4 variabili latenti di primo ordine con 5 manifeste per blocco

La seconda simulazione è stata effettuata considerando un modello con 4 VL di primo ordine composte ognuna da 5 VM.

5.3.1 Confronto comunalità

La tabella 5.10 riporta i risultati relativi alla comunalità, alle bias e gli errori standard raggruppati in base all'approccio, alle relazioni del modello esterno ed interno e alla numerosità del campione.

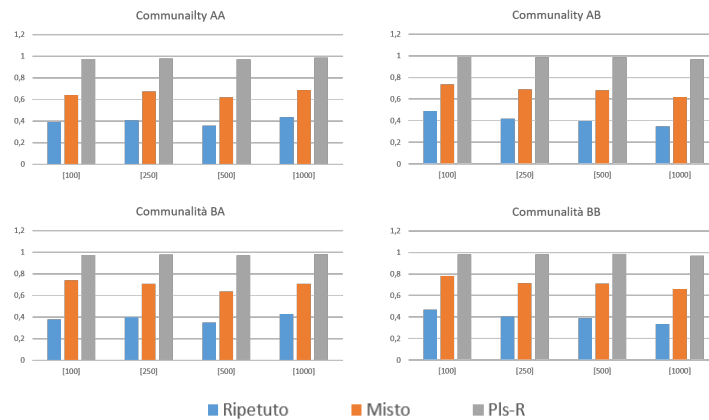


Fig. 5.3: Confronto Comunalità

La comunalità stimata è sempre significativa. Considerando il valore dell'indicatore, come si può notare in figura 5.3, esso è sempre più alto utilizzando l'algoritmo del PLS-R, sia all'aumentare della numerosità del campione sia con ogni tipologia di rapporto del modello. In tabella 5.10 si nota inoltre come all'aumentare del numero del campione la variabilità della stima sia inferiore per il PLS-R rispetto gli altri approcci.

5.3.2 Path Coefficient

Le seguenti tabelle riportano i risultati relativi ai coefficienti β , alle bias e gli errori standard raggruppati in base all'approccio, alle relazioni del modello esterno ed interno e alla numerosità del campione.

Come si può notare nelle tabelle 5.11, 5.12, 5.13, 5.14, al variare della numerosità campionaria e dell'approccio utilizzato, i valori dei vari path non

Tab. 5.10: Confronto Comunalità

Approccio	Numerosità	Valore	Rif.-Form.	Form.-Rif.	Rif.-Rif.	Form.-Form.
Ripetuto	100	Path	0,49119	0,37671	0,38944	0,46839
		Bias	0,00018	-0,00239	0,00719	-0,00225
		Errore St.	0,03100	0,03491	0,03786	0,03848
	250	Path	0,41699	0,39962	0,40902	0,40648
		Bias	0,00076	-0,00198	0,00178	-0,00110
		Errore St.	0,02110	0,02325	0,01998	0,02135
	500	Path	0,39658	0,34819	0,35774	0,38896
		Bias	0,00072	-0,00147	0,00059	-0,00004
		Errore St.	0,01343	0,01451	0,01348	0,01360
	1000	Path	0,34624	0,42947	0,43868	0,33471
		Bias	0,00113	-0,00009	0,00018	-0,00015
		Errore St.	0,01052	0,01013	0,01037	0,00954
Misto	100	Path	0,73705	0,73912	0,63957	0,78304
		Bias	0,00182	0,01934	0,00900	0,01477
		Errore St.	0,02625	0,02492	0,03897	0,02238
	250	Path	0,68808	0,70729	0,67608	0,71354
		Bias	0,00232	0,00726	0,00188	0,00699
		Errore St.	0,02110	0,02225	0,01962	0,01914
	500	Path	0,68100	0,63841	0,61826	0,70918
		Bias	0,00091	0,00456	0,00084	0,00294
		Errore St.	0,01337	0,01632	0,01514	0,01328
	1000	Path	0,61806	0,70984	0,68758	0,65937
		Bias	0,00148	0,00206	0,00073	0,00279
		Errore St.	0,01203	0,00974	0,01042	0,01093
PLS-R	100	Path	0,99200	0,97429	0,97349	0,98348
		Bias	-0,01070	-0,17005	-0,01572	-0,01892
		Errore St.	0,04084	0,09023	0,03669	0,04569
	250	Path	0,98721	0,97716	0,97976	0,98038
		Bias	-0,00163	-0,00707	-0,00711	-0,00241
		Errore St.	0,00450	0,02636	0,02711	0,00604
	500	Path	0,98549	0,96961	0,97063	0,98499
		Bias	-0,00071	-0,00213	-0,00151	-0,00103
		Errore St.	0,00283	0,00646	0,00592	0,00296
	1000	Path	0,96828	0,98299	0,98566	0,96768
		Bias	-0,00046	-0,00036	-0,00030	-0,00082
		Errore St.	0,00492	0,00231	0,00226	0,00452

si discostano molto tra loro. Le stime dei path sono sempre significative ma, pur rimanendo bassa, la variabilità delle stime per l'approccio PLS-R è più alta rispetto a quella degli altri approcci.

Tab. 5.11: Modello Formativo-Formativo

Approccio	Numerosità	Valore	β_1	β_2	β_3	β_4
Ripetuto	100	Path	0,28288	0,28285	0,28284	0,28283
		Bias	-0,00268	-0,00268	-0,00266	-0,00265
		Errore St.	0,00410	0,00410	0,00410	0,00411
	250	Path	0,29587	0,29591	0,29596	0,29588
		Bias	-0,00125	-0,00126	-0,00128	-0,00126
		Errore St.	0,00394	0,00394	0,00394	0,00394
	500	Path	0,29697	0,29696	0,29696	0,29699
		Bias	-0,00066	-0,00066	-0,00065	-0,00067
		Errore St.	0,00277	0,00277	0,00277	0,00277
	1000	Path	0,30792	0,30791	0,30793	0,30793
		Bias	-0,00065	-0,00065	-0,00064	-0,00064
		Errore St.	0,00254	0,00254	0,00254	0,00254
Misto	100	Path	0,28276	0,28287	0,28289	0,28285
		Bias	-0,00260	-0,00268	-0,00270	-0,00267
		Errore St.	0,00410	0,00410	0,00411	0,00411
	250	Path	0,29590	0,29586	0,29597	0,29583
		Bias	-0,00128	-0,00122	-0,00131	-0,00122
		Errore St.	0,00394	0,00395	0,00394	0,00394
	500	Path	0,29698	0,29698	0,29691	0,29695
		Bias	-0,00066	-0,00067	-0,00065	-0,00066
		Errore St.	0,00277	0,00277	0,00277	0,00277
	1000	Path	0,30797	0,30794	0,30793	0,30777
		Bias	-0,00065	-0,00065	-0,00065	-0,00061
		Errore St.	0,00254	0,00254	0,00254	0,00255
PLS-R	100	Path	0,25463	0,31544	0,29234	0,26933
		Bias	0,00067	-0,00880	0,00237	-0,00596
		Errore St.	0,02996	0,01616	0,02519	0,03095
	250	Path	0,31195	0,30891	0,25835	0,30332
		Bias	-0,00246	-0,00044	0,00031	-0,00273
		Errore St.	0,01035	0,01191	0,01468	0,01141
	500	Path	0,28795	0,29363	0,30134	0,30256
		Bias	-0,00111	-0,00089	-0,00046	-0,00035
		Errore St.	0,00867	0,00906	0,01039	0,00920
	1000	Path	0,31584	0,33597	0,30515	0,27138
		Bias	-0,00205	-0,00034	0,00080	-0,00109
		Errore St.	0,00713	0,00614	0,00898	0,00973

Tab. 5.12: Modello Riflessivo-Riflessivo

Approccio	Numerosità	Valore	β_1	β_2	β_3	β_4
Ripetuto	100	Path	0,75836	0,82322	0,81044	0,80664
		Bias	-0,00038	0,00590	0,00407	0,00724
		Errore St.	0,05101	0,02959	0,03223	0,03909
	250	Path	0,71086	0,84437	0,84534	0,87814
		Bias	-0,00032	0,00194	0,00066	0,00089
		Errore St.	0,03699	0,01625	0,01824	0,01401
	500	Path	0,78861	0,82010	0,74706	0,78512
		Bias	0,00098	0,00015	0,00184	0,00018
		Errore St.	0,01801	0,01514	0,02088	0,01916
	1000	Path	0,87577	0,87977	0,72944	0,82223
		Bias	-0,00007	0,00015	0,00095	0,00047
		Errore St.	0,00681	0,00717	0,01580	0,01075
Misto	100	Path	0,76134	0,82370	0,81018	0,80345
		Bias	-0,00042	0,00595	0,00430	0,00776
		Errore St.	0,04904	0,02871	0,03070	0,03557
	250	Path	0,72385	0,84308	0,83460	0,87920
		Bias	0,00064	0,00139	0,00095	0,00058
		Errore St.	0,03327	0,01547	0,01804	0,01383
	500	Path	0,78139	0,81140	0,76507	0,78472
		Bias	0,00157	-0,00022	0,00158	0,00044
		Errore St.	0,01740	0,01498	0,01849	0,01785
	1000	Path	0,86814	0,86875	0,74782	0,82642
		Bias	-0,00021	0,00023	0,00100	0,00051
		Errore St.	0,00709	0,00757	0,01427	0,00993
PLS-R	100	Path	0,76084	0,82273	0,80996	0,80518
		Bias	-0,00012	0,00552	0,00297	0,00778
		Errore St.	0,04872	0,02936	0,03466	0,03894
	250	Path	0,71665	0,84258	0,84307	0,87742
		Bias	0,00142	0,00050	0,00016	0,00101
		Errore St.	0,03525	0,01765	0,01828	0,01419
	500	Path	0,78776	0,81869	0,74926	0,78542
		Bias	0,00115	0,00026	0,00176	0,00002
		Errore St.	0,01760	0,01501	0,01982	0,01851
	1000	Path	0,87428	0,87865	0,73369	0,82144
		Bias	-0,00014	0,00007	0,00111	0,00049
		Errore St.	0,00693	0,00726	0,01481	0,01062

Tab. 5.13: Modello Formativo-Riflessivo

Approccio	Numerosità	Valore	β_1	β_2	β_3	β_4
Ripetuto	100	Path	0,86702	0,86757	0,86186	0,84200
		Bias	0,01182	0,01262	0,00737	0,01231
		Errore St.	0,01572	0,02132	0,02420	0,02537
	250	Path	0,77563	0,85845	0,85780	0,86863
		Bias	0,00780	0,00398	0,00332	0,00204
		Errore St.	0,02501	0,01697	0,01641	0,01625
	500	Path	0,78818	0,82176	0,77303	0,81206
		Bias	0,00361	0,00245	0,00342	0,00184
		Errore St.	0,01650	0,01312	0,01727	0,01461
	1000	Path	0,86940	0,87705	0,78452	0,83618
		Bias	0,00101	0,00039	0,00189	0,00163
		Errore St.	0,00725	0,00662	0,01163	0,00861
Misto	100	Path	0,86695	0,86743	0,86192	0,84204
		Bias	0,01186	0,01274	0,00730	0,01225
		Errore St.	0,01573	0,02131	0,02423	0,02538
	250	Path	0,77563	0,85844	0,85773	0,86857
		Bias	0,00778	0,00396	0,00337	0,00207
		Errore St.	0,02502	0,01697	0,01641	0,01625
	500	Path	0,78824	0,82167	0,77293	0,81198
		Bias	0,00351	0,00251	0,00346	0,00187
		Errore St.	0,01651	0,01311	0,01727	0,01462
	1000	Path	0,86936	0,87700	0,78446	0,83614
		Bias	0,00101	0,00037	0,00191	0,00164
		Errore St.	0,00725	0,00662	0,01163	0,00861
PLS-R	100	Path	0,84764	0,86502	0,83798	0,83774
		Bias	0,01387	0,01306	0,01107	0,01023
		Errore St.	0,02194	0,02329	0,03164	0,02657
	250	Path	0,76355	0,84854	0,85281	0,87571
		Bias	0,00385	0,00424	0,00175	0,00198
		Errore St.	0,02818	0,01774	0,01798	0,01542
	500	Path	0,79089	0,81526	0,77219	0,79453
		Bias	0,00214	0,00177	0,00124	0,00240
		Errore St.	0,01691	0,01411	0,01779	0,01564
	1000	Path	0,86968	0,87083	0,77196	0,83645
		Bias	0,00089	0,00037	0,00105	0,00143
		Errore St.	0,00748	0,00720	0,01183	0,00896

Tab. 5.14: Modello Riflessivo-Formativo

Approccio	Numerosità	Valore	β_1	β_2	β_3	β_4
Ripetuto	100	Path	0,32566	0,23644	0,30515	0,29293
		Bias	-0,00059	0,00338	-0,00140	-0,00242
		Errore St.	0,01459	0,01577	0,01253	0,01399
	250	Path	0,25419	0,34305	0,31544	0,28867
		Bias	-0,00014	-0,00007	-0,00083	-0,00030
		Errore St.	0,01075	0,01289	0,01136	0,01265
	500	Path	0,28927	0,31094	0,32024	0,29079
		Bias	-0,00038	-0,00050	0,00025	-0,00020
		Errore St.	0,00839	0,00822	0,00928	0,00849
	1000	Path	0,30708	0,33581	0,33454	0,29379
		Bias	-0,00078	-0,00019	-0,00064	-0,00016
		Errore St.	0,00757	0,00776	0,00730	0,00822
Misto	100	Path	0,29698	0,26975	0,30229	0,29504
		Bias	-0,00049	0,00214	-0,00318	0,00032
		Errore St.	0,00800	0,00740	0,00823	0,00794
	250	Path	0,29303	0,30909	0,30417	0,29815
		Bias	-0,00095	0,00083	-0,00014	-0,00092
		Errore St.	0,00618	0,00760	0,00695	0,00680
	500	Path	0,29861	0,30557	0,30445	0,30306
		Bias	-0,00015	0,00031	0,00043	-0,00131
		Errore St.	0,00446	0,00431	0,00486	0,00431
	1000	Path	0,31808	0,31579	0,32703	0,31110
		Bias	-0,00036	-0,00116	-0,00036	0,00023
		Errore St.	0,00472	0,00431	0,00462	0,00445
PLS-R	100	Path	0,32356	0,24273	0,30294	0,29172
		Bias	0,00051	0,00173	-0,00017	-0,00381
		Errore St.	0,02306	0,02172	0,01757	0,02673
	250	Path	0,25678	0,34119	0,31554	0,28801
		Bias	0,00004	-0,00021	-0,00150	0,00049
		Errore St.	0,00989	0,01219	0,01078	0,01167
	500	Path	0,29122	0,31068	0,31879	0,29060
		Bias	-0,00105	-0,00060	0,00063	0,00021
		Errore St.	0,00779	0,00780	0,00867	0,00793
	1000	Path	0,30774	0,33353	0,33275	0,29741
		Bias	-0,00094	0,00030	-0,00097	-0,00009
		Errore St.	0,00686	0,00694	0,00668	0,00727

Tab. 5.15: Relative Bias Modello Formativo-Formativo

Approccio	Path	100	250	500	1000
Ripetuto	β_1	-0,00949	-0,00422	-0,00222	-0,00210
	β_2	-0,00947	-0,00427	-0,00221	-0,00210
	β_3	-0,00940	-0,00431	-0,00220	-0,00209
	β_4	-0,00936	-0,00426	-0,00225	-0,00209
Misto	β_1	-0,00921	-0,00433	-0,00221	-0,00212
	β_2	-0,00949	-0,00414	-0,00225	-0,00211
	β_3	-0,00954	-0,00443	-0,00220	-0,00212
	β_4	-0,00945	-0,00414	-0,00221	-0,00199
PLS-R	β_1	0,00261	-0,00788	-0,00386	-0,00650
	β_2	-0,02790	-0,00143	-0,00302	-0,00101
	β_3	0,00811	0,00118	-0,00152	0,00263
	β_4	-0,02214	-0,00901	-0,00114	-0,00403

5.3.3 Relative Bias

Nella Tabella 5.15 troviamo i valori degli RB per i path del modello Formativo-Formativo.

Osservando gli RB si nota che essi sono sempre molto vicino allo zero, dunque in tutti gli approcci i path risultano ben stimati, per l'approccio PLS-R però gli RB sono quasi sempre maggiori in valore assoluto, e quindi più distanti dal valore effettivo del parametro. Nella Tabella 5.16 troviamo i valori degli RB per i path del modello Riflessivo-Riflessivo.

In questo caso gli RB sono sempre molto simili tra loro, e in generale molto vicini allo zero, quindi anche in questo caso i *path*, per tutti gli approcci, risultano ben stimati. Nella Tabella 5.17 troviamo i valori degli RB per i *path* del modello Riflessivo-Formativo.

Nel caso dello schema misto Riflessivo-Formativo, si nota come per i parametri β_2 e β_4 il PLS-R dia stime migliori all'aumentare del numero di osservazioni, mentre per gli altri *path* i risultati siano altalenanti, anche se comunque sempre molto vicini allo zero. Nella Tabella 5.18 troviamo i valori degli RB per i *path* del modello Formativo-Riflessivo.

In questo caso le stime per il PLS-R sono sempre in valore assoluto più basse, a parte per i β_1 e β_3 , e tendono a decrescere con l'aumentare del numero di osservazioni.

Tab. 5.16: Relative Bias Modello Riflessivo-Riflessivo

Approccio	Path	100	250	500	1000
Ripetuto	β_1	-0,00050	-0,00045	0,00124	-0,00008
	β_2	0,00717	0,00229	0,00018	0,00017
	β_3	0,00502	0,00078	0,00246	0,00130
	β_4	0,00897	0,00102	0,00023	0,00058
Misto	β_1	-0,00056	0,00088	0,00201	-0,00024
	β_2	0,00723	0,00164	-0,00028	0,00026
	β_3	0,00531	0,00114	0,00206	0,00134
	β_4	0,00966	0,00066	0,00056	0,00061
PLS-R	β_1	-0,00016	0,00198	0,00146	-0,00016
	β_2	0,00671	0,00059	0,00031	0,00008
	β_3	0,00367	0,00019	0,00235	0,00152
	β_4	0,00967	0,00116	0,00003	0,00059

Tab. 5.17: Relative Bias Modello Riflessivo-Formativo

Approccio	Path	100	250	500	1000
Ripetuto	β_1	-0,00183	-0,00054	-0,00132	-0,00253
	β_2	0,01428	-0,00020	-0,00161	-0,00057
	β_3	-0,00460	-0,00262	0,00078	-0,00192
	β_4	-0,00826	-0,00103	-0,00068	-0,00054
Misto	β_1	-0,00165	-0,00326	-0,00049	-0,00114
	β_2	0,00793	0,00270	0,00102	-0,00368
	β_3	-0,01051	-0,00045	0,00141	-0,00111
	β_4	0,00109	-0,00308	-0,00431	0,00072
PLS-R	β_1	0,00156	0,00018	-0,00361	-0,00307
	β_2	0,00713	-0,00062	-0,00193	0,00091
	β_3	-0,00055	-0,00476	0,00198	-0,00291
	β_4	-0,01307	0,00169	0,00074	-0,00029

<i>Tab. 5.18: Relative Bias Modello Formativo-Riflessivo</i>					
Approccio	Path	100	250	500	1000
Ripetuto	β_1	0,01363	0,01005	0,00458	0,00116
	β_2	0,01455	0,00464	0,00299	0,00044
	β_3	0,00855	0,00387	0,00443	0,00240
	β_4	0,01462	0,00235	0,00226	0,00195
Misto	β_1	0,01368	0,01003	0,00446	0,00116
	β_2	0,01469	0,00461	0,00306	0,00042
	β_3	0,00847	0,00392	0,00447	0,00243
	β_4	0,01455	0,00238	0,00230	0,00196
PLS-R	β_1	0,01637	0,00504	0,00271	0,00102
	β_2	0,01509	0,00500	0,00217	0,00043
	β_3	0,01321	0,00205	0,00160	0,00137
	β_4	0,01221	0,00226	0,00302	0,00171

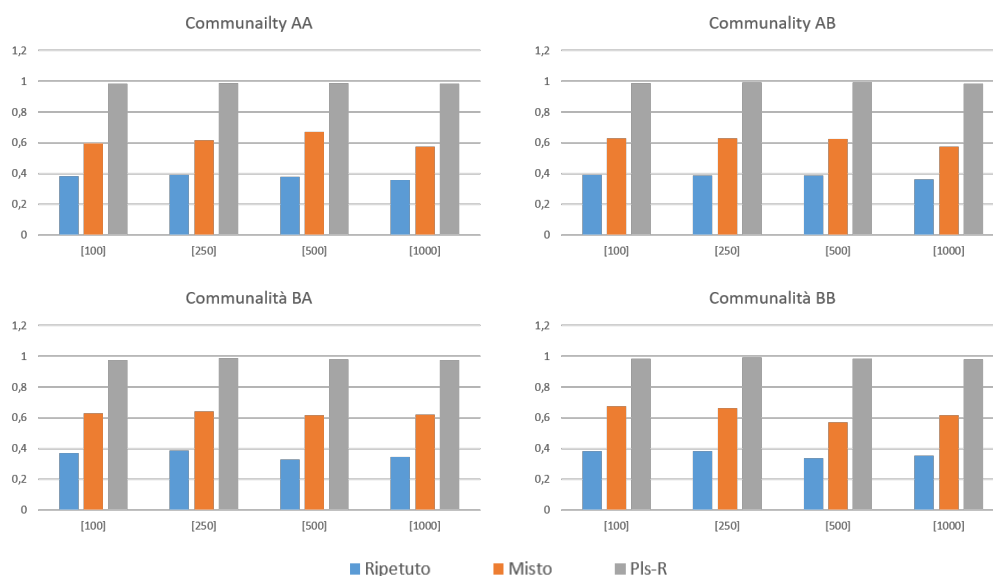


Fig. 5.4: Confronto Comunalità

5.4 Risultati: Modello a 7 variabili latenti di primo ordine con 4 manifeste per blocco

L'ultima simulazione è stata effettuata considerando un modello con 7 VL di primo ordine composte ognuna da 4 VM.

5.4.1 Confronto comunalità

La tabella 5.19 riporta i risultati relativi alla comunalità, alle bias e gli errori standard raggruppati in base all'approccio, alle relazioni del modello esterno ed interno e alla numerosità del campione.

La comunalità stimata è sempre significativa. Considerando il valore dell'indicatore, come si può notare in figura 5.4, esso è sempre più alto utilizzando l'algoritmo del PLS-R, sia all'aumentare della numerosità del campione sia con ogni tipologia di rapporto del modello. In tabella 5.19 si nota inoltre come all'aumentare del numero del campione la variabilità della stima sia inferiore per il PLS-R rispetto a quando si utilizzano gli altri approcci.

Tab. 5.19: Confronto Comunalità

Approccio	Numerosità	Valore	Rif.-Form.	From.-Rif.	Rif.-Rif.	From.-From.
Ripetuto	100	Path	0,39185	0,37152	0,38217	0,38449
		Bias	0,00104	-0,00495	-0,00127	-0,00574
		Errore St.	0,03026	0,03104	0,02983	0,03512
	250	Path	0,38835	0,38565	0,39274	0,38233
		Bias	0,00126	0,00044	-0,00054	-0,00151
		Errore St.	0,02354	0,02081	0,02128	0,02244
	500	Path	0,38732	0,33056	0,38054	0,33901
		Bias	0,00087	-0,00061	0,00077	0,00120
		Errore St.	0,01323	0,01251	0,01331	0,01155
	1000	Path	0,36074	0,34732	0,35882	0,35257
		Bias	-0,00020	0,00007	0,00033	0,00016
		Errore St.	0,00963	0,00914	0,00965	0,01010
Misto	100	Path	0,62829	0,63097	0,59485	0,67636
		Bias	0,00442	0,00861	0,00085	0,00542
		Errore St.	0,03222	0,03702	0,03522	0,03464
	250	Path	0,62970	0,64003	0,61736	0,66072
		Bias	0,00238	0,00735	0,00064	0,00408
		Errore St.	0,02475	0,02096	0,02201	0,02301
	500	Path	0,62632	0,61484	0,67172	0,57232
		Bias	0,00193	0,00240	0,00303	0,00214
		Errore St.	0,01429	0,01438	0,01430	0,01317
	1000	Path	0,57669	0,62020	0,57318	0,61826
		Bias	-0,00016	0,00127	0,00064	0,00170
		Errore St.	0,01111	0,01026	0,01096	0,01116
PLS-R	100	Path	0,98560	0,97702	0,98258	0,98212
		Bias	-0,02980	-0,01265	-0,01024	-0,03577
		Errore St.	0,03872	0,01475	0,01972	0,03867
	250	Path	0,99156	0,98669	0,98956	0,99010
		Bias	-0,00152	-0,00240	-0,00171	-0,00280
		Errore St.	0,00332	0,00450	0,00342	0,00365
	500	Path	0,99025	0,97906	0,98747	0,98360
		Bias	-0,00082	-0,00234	-0,00111	-0,00188
		Errore St.	0,00413	0,00622	0,00250	0,00744
	1000	Path	0,98480	0,97650	0,98297	0,97866
		Bias	-0,00074	-0,00080	-0,00059	-0,00088
		Errore St.	0,00218	0,00290	0,00237	0,00268

Tab. 5.20: Modello Formativo-Formativo

Approccio	Numerosità	Valore	β_1	β_2	β_3	β_4	β_5	β_6	β_7
Ripetuto	100	Path	0,79085	0,68325	0,66318	0,75119	0,87336	0,82399	0,78808
		Bias	0,00235	0,00268	-0,00970	0,00216	0,00143	-0,00458	0,00050
		Errore St.	0,03404	0,05076	0,06162	0,04972	0,02034	0,02773	0,03714
	250	Path	0,82980	0,70160	0,77809	0,83068	0,75373	0,80914	0,78768
		Bias	0,00003	0,00075	0,00050	0,00037	-0,00079	-0,00021	0,00017
		Errore St.	0,01936	0,03609	0,02465	0,02095	0,02846	0,02271	0,02483
	500	Path	0,77446	0,77452	0,77453	0,77450	0,74511	0,74464	0,77452
		Bias	-0,00038	-0,00038	-0,00039	-0,00038	-0,00039	-0,00038	-0,00039
		Errore St.	0,00187	0,00187	0,00187	0,00187	0,00187	0,00187	0,00187
	1000	Path	0,75663	0,66514	0,73395	0,68681	0,83915	0,79831	0,80111
		Bias	0,00045	0,00065	0,00022	0,00067	-0,00027	0,00065	0,00054
		Errore St.	0,01445	0,01787	0,01434	0,01758	0,00868	0,01142	0,01187
Misto	100	Path	0,79475	0,70114	0,66328	0,74643	0,85816	0,82560	0,79113
		Bias	0,00200	0,00147	-0,00981	0,00280	0,00150	-0,00441	0,00137
		Errore St.	0,03280	0,04868	0,06177	0,04785	0,02237	0,02720	0,03616
	250	Path	0,82189	0,70547	0,78167	0,83097	0,75828	0,81273	0,78129
		Bias	0,00002	-0,00006	0,00060	0,00060	-0,00073	0,00017	0,00046
		Errore St.	0,02012	0,03490	0,02368	0,02010	0,02761	0,02144	0,02482
	500	Path	0,77448	0,77453	0,77449	0,77449	0,77448	0,77452	0,77448
		Bias	-0,00040	-0,00039	-0,00037	-0,00036	-0,00038	-0,00039	-0,00038
		Errore St.	0,00187	0,00187	0,00187	0,00187	0,00187	0,00187	0,00187
	1000	Path	0,75304	0,67341	0,73866	0,70233	0,83220	0,78908	0,79766
		Bias	0,00047	0,00086	0,00021	0,00038	-0,00021	0,00092	0,00033
		Errore St.	0,01428	0,01696	0,01393	0,01677	0,00881	0,01158	0,01168
PLS-R	100	Path	0,78981	0,68455	0,66487	0,75153	0,87280	0,82323	0,78779
		Bias	0,00178	0,00309	-0,00923	0,00240	0,00074	-0,00404	0,00016
		Errore St.	0,03409	0,04933	0,05866	0,04909	0,02076	0,02811	0,03720
	250	Path	0,82914	0,70310	0,77797	0,82962	0,75419	0,80895	0,78794
		Bias	0,00010	0,00095	0,00059	0,00074	-0,00114	-0,00033	-0,00002
		Errore St.	0,01938	0,03512	0,02442	0,02093	0,02810	0,02262	0,02460
	500	Path	0,74797	0,79942	0,76776	0,77502	0,75905	0,78111	0,78738
		Bias	0,00235	-0,00034	-0,00052	-0,00114	-0,00252	-0,00153	0,00070
		Errore St.	0,00938	0,00824	0,00856	0,00868	0,00913	0,00771	0,00500
	1000	Path	0,75650	0,66750	0,73402	0,68785	0,83857	0,79664	0,80066
		Bias	0,00035	0,00053	0,00019	0,00078	-0,00027	0,00082	0,00051
		Errore St.	0,01429	0,01723	0,01415	0,01714	0,00869	0,01145	0,01184

5.4.2 Path Coefficient

Le seguenti tabelle riportano i risultati relativi ai coefficienti β , alle bias e gli errori standard raggruppati in base all'approccio, alle relazioni del modello esterno ed interno e alla numerosità del campione.

Come si può notare nelle tabelle 5.20, 5.21, 5.22, 5.23, al variare della numerosità campionaria e dell'approccio utilizzato, i valori dei vari path non si discostano molto tra loro. Le stime dei path sono sempre significative, ma pur rimanendo bassa, la variabilità delle stime per l'approccio PLS-R è più alta rispetto a quella degli altri approcci.

Tab. 5.21: Modello Riflessivo-Riflessivo

Approccio	Numerosità	Valore	β_1	β_2	β_3	β_4	β_5	β_6	β_7
Ripetuto	100	Path	0,17438	0,17444	0,17439	0,17440	0,17440	0,17442	0,17439
		Bias	-0,00047	-0,00048	-0,00047	-0,00047	-0,00048	-0,00048	-0,00048
		Errore St.	0,00469	0,00469	0,00469	0,00470	0,00469	0,00469	0,00469
	250	Path	0,17582	0,17583	0,17588	0,17584	0,17586	0,17584	0,17583
		Bias	-0,00039	-0,00040	-0,00042	-0,00040	-0,00041	-0,00040	-0,00039
		Errore St.	0,00308	0,00308	0,00308	0,00308	0,00308	0,00308	0,00308
	500	Path	0,07286	0,20755	0,15891	0,13491	0,20216	0,17525	0,13533
		Bias	0,00114	0,00046	0,00265	0,00177	0,00165	-0,00055	0,00090
		Errore St.	0,02643	0,01469	0,01930	0,01894	0,01580	0,01745	0,02071
	1000	Path	0,18228	0,18228	0,18229	0,18232	0,18229	0,18229	0,18230
		Bias	-0,00022	-0,00022	-0,00022	-0,00022	-0,00022	-0,00022	-0,00021
		Errore St.	0,00169	0,00169	0,00169	0,00169	0,00169	0,00169	0,00169
Misto	100	Path	0,17435	0,17445	0,17442	0,17433	0,17435	0,17443	0,17446
		Bias	-0,00045	-0,00050	-0,00047	-0,00042	-0,00044	-0,00049	-0,00054
		Errore St.	0,00469	0,00469	0,00468	0,00470	0,00469	0,00469	0,00469
	250	Path	0,17578	0,17581	0,17588	0,17585	0,17584	0,17584	0,17586
		Bias	-0,00035	-0,00041	-0,00042	-0,00041	-0,00040	-0,00040	-0,00042
		Errore St.	0,00308	0,00308	0,00308	0,00308	0,00308	0,00308	0,00308
	500	Path	0,18338	0,20095	0,15900	0,13817	0,20297	0,17293	0,13136
		Bias	0,00120	0,00019	0,00244	0,00163	0,00165	-0,00020	0,00136
		Errore St.	0,02510	0,01451	0,01893	0,01806	0,01501	0,01697	0,02040
	1000	Path	0,18227	0,18229	0,18226	0,18235	0,18227	0,18228	0,18229
		Bias	-0,00023	-0,00023	-0,00021	-0,00024	-0,00020	-0,00022	-0,00019
		Errore St.	0,00169	0,00169	0,00169	0,00169	0,00169	0,00169	0,00169
PLS-R	100	Path	0,16480	0,18131	0,17968	0,15529	0,17711	0,19571	0,16422
		Bias	0,00438	0,00121	-0,00298	0,00681	-0,00842	-0,00642	-0,00085
		Errore St.	0,01948	0,02347	0,02048	0,02777	0,01589	0,01608	0,02290
	250	Path	0,16933	0,17613	0,15298	0,18932	0,17322	0,17241	0,19453
		Bias	-0,00025	0,00095	-0,00158	0,00027	-0,00004	0,00014	-0,00248
		Errore St.	0,01230	0,01134	0,01380	0,01153	0,01101	0,01118	0,01174
	500	Path	0,07461	0,20659	0,15880	0,13536	0,20152	0,17476	0,13560
		Bias	0,00144	0,00057	0,00242	0,00126	0,00187	-0,00040	0,00080
		Errore St.	0,02578	0,01471	0,01912	0,01926	0,01579	0,01737	0,02037
	1000	Path	0,17978	0,19701	0,19480	0,14631	0,19614	0,19690	0,16831
		Bias	-0,00117	-0,00103	-0,00089	-0,00059	-0,00010	0,00065	0,00119
		Errore St.	0,00591	0,00419	0,00292	0,00736	0,00457	0,00508	0,00797

Tab. 5.22: Modello Formativo-Riflessivo

Approccio	Numerosità	Valore	β_1	β_2	β_3	β_4	β_5	β_6	β_7
Ripetuto	100	Path	0,80622	0,74002	0,69987	0,78184	0,86518	0,85521	0,80000
		Bias	0,00387	0,00528	0,00261	0,00705	0,00019	0,00458	0,00791
		Errore St.	0,03528	0,03902	0,04822	0,03861	0,02436	0,02382	0,03989
	250	Path	0,82502	0,71689	0,81058	0,84315	0,77169	0,83322	0,79627
		Bias	0,00318	0,00644	0,00412	0,00299	0,00417	0,00188	0,00408
		Errore St.	0,02022	0,02882	0,02078	0,01964	0,02542	0,01936	0,02003
	500	Path	0,72056	0,82469	0,76526	0,82310	0,82936	0,77211	0,74712
		Bias	0,00328	0,00103	0,00134	0,00064	0,00112	0,00083	0,00173
		Errore St.	0,01987	0,01378	0,01861	0,01398	0,01302	0,01670	0,01862
	1000	Path	0,80120	0,73954	0,74392	0,74400	0,84743	0,81846	0,81181
		Bias	0,00017	0,00124	0,00111	0,00081	0,00051	0,00103	0,00000
		Errore St.	0,01054	0,01358	0,01310	0,01336	0,00868	0,01025	0,00979
Misto	100	Path	0,80621	0,74010	0,69983	0,78178	0,86505	0,85529	0,79997
		Bias	0,00385	0,00520	0,00266	0,00710	0,00030	0,00449	0,00791
		Errore St.	0,03527	0,03902	0,04822	0,03860	0,02437	0,02382	0,03991
	250	Path	0,82502	0,71680	0,81061	0,84310	0,77174	0,83307	0,79632
		Bias	0,00315	0,00649	0,00408	0,00302	0,00411	0,00200	0,00402
		Errore St.	0,02022	0,02883	0,02078	0,01964	0,02542	0,01936	0,02003
	500	Path	0,72052	0,82468	0,76526	0,82304	0,82933	0,77207	0,74711
		Bias	0,00330	0,00102	0,00132	0,00067	0,00112	0,00083	0,00173
		Errore St.	0,01987	0,01378	0,01861	0,01398	0,01303	0,01670	0,01862
	1000	Path	0,80117	0,73952	0,74388	0,74396	0,84740	0,81842	0,81179
		Bias	0,00018	0,00123	0,00112	0,00082	0,00050	0,00104	-0,00001
		Errore St.	0,01054	0,01358	0,01310	0,01336	0,00868	0,01026	0,00979
PLS-R	100	Path	0,79872	0,73240	0,69013	0,77363	0,87044	0,84112	0,79983
		Bias	0,00314	0,00380	-0,00066	0,00611	-0,00103	0,00743	0,00752
		Errore St.	0,03743	0,04132	0,04859	0,04363	0,02632	0,02600	0,03836
	250	Path	0,82015	0,71875	0,79833	0,83839	0,77213	0,82369	0,79873
		Bias	0,00393	0,00422	0,00461	0,00185	0,00201	0,00250	0,00433
		Errore St.	0,02100	0,02947	0,02162	0,01970	0,02622	0,02030	0,02081
	500	Path	0,71540	0,81907	0,77090	0,78843	0,82980	0,77776	0,73862
		Bias	0,00296	0,00090	0,00017	0,00142	0,00107	-0,00014	0,00192
		Errore St.	0,02082	0,01439	0,01941	0,01663	0,01371	0,01711	0,01900
	1000	Path	0,78295	0,70850	0,75065	0,73894	0,84242	0,81705	0,81293
		Bias	-0,00019	0,00107	0,00093	0,00033	0,00096	0,00097	0,00011
		Errore St.	0,01116	0,01479	0,01342	0,01384	0,00906	0,01068	0,01017

Tab. 5.23: Modello Riflessivo-Formativo

Approccio	Numerosità	Valore	β_1	β_2	β_3	β_4	β_5	β_6	β_7
Ripetuto	100	Path	0,17743	0,19009	0,22108	0,18738	0,13540	0,18098	0,16016
		Bias	-0,00065	-0,00288	-0,00182	0,00275	-0,00161	-0,00134	0,00073
		Errore St.	0,00957	0,01336	0,01168	0,01714	0,01201	0,01014	0,01966
	250	Path	0,17472	0,18165	0,19123	0,18516	0,16210	0,14737	0,21371
		Bias	0,00019	-0,00061	-0,00075	-0,00058	0,00029	-0,00038	-0,00022
		Errore St.	0,00722	0,00835	0,00854	0,00812	0,00884	0,00806	0,00809
	500	Path	0,20383	0,22343	0,16889	0,16725	0,15370	0,17680	0,16271
		Bias	-0,00037	-0,00037	-0,00046	-0,00015	0,00037	-0,00026	-0,00034
		Errore St.	0,00597	0,00542	0,00552	0,00554	0,00593	0,00553	0,00574
	1000	Path	0,16767	0,19394	0,14346	0,21521	0,15859	0,18385	0,24180
		Bias	-0,00023	0,00000	-0,00012	-0,00024	0,00040	-0,00010	0,00019
		Errore St.	0,00452	0,00433	0,00472	0,00530	0,00462	0,00433	0,00505
Misto	100	Path	0,18712	0,19102	0,19481	0,18394	0,14847	0,18193	0,17058
		Bias	-0,00131	-0,00076	-0,00067	0,00004	-0,00132	-0,00083	0,00051
		Errore St.	0,00730	0,00774	0,00818	0,00760	0,00868	0,00579	0,01156
	250	Path	0,18268	0,18356	0,18751	0,17899	0,17505	0,16503	0,18620
		Bias	0,00004	-0,00066	-0,00111	-0,00047	0,00061	0,00028	-0,00045
		Errore St.	0,00508	0,00553	0,00580	0,00513	0,00554	0,00553	0,00532
	500	Path	0,18939	0,19444	0,18148	0,17872	0,16588	0,17760	0,17405
		Bias	-0,00061	-0,00027	0,00029	-0,00055	0,00009	0,00017	-0,00065
		Errore St.	0,00357	0,00353	0,00344	0,00364	0,00376	0,00345	0,00384
	1000	Path	0,18113	0,18580	0,16879	0,19697	0,18321	0,18956	0,20883
		Bias	0,00008	0,00075	-0,00032	-0,00011	0,00034	-0,00055	-0,00018
		Errore St.	0,00321	0,00283	0,00365	0,00330	0,00306	0,00308	0,00340
PLS-R	100	Path	0,17898	0,18821	0,21988	0,18584	0,14178	0,18075	0,15853
		Bias	-0,00001	0,00172	0,00530	-0,01552	-0,00027	0,00093	0,00232
		Errore St.	0,01129	0,02299	0,01703	0,04443	0,01066	0,01081	0,02149
	250	Path	0,17444	0,17993	0,19164	0,18560	0,16305	0,14943	0,21216
		Bias	0,00021	0,00055	-0,00184	-0,00081	-0,00036	0,00003	0,00031
		Errore St.	0,00703	0,00794	0,00821	0,00779	0,00843	0,00742	0,00779
	500	Path	0,20206	0,22208	0,16958	0,16688	0,15640	0,17698	0,16310
		Bias	-0,00008	-0,00027	-0,00084	0,00030	0,00000	-0,00075	0,00007
		Errore St.	0,00731	0,00540	0,00534	0,00531	0,00542	0,00532	0,00544
	1000	Path	0,16890	0,19279	0,14759	0,21328	0,15992	0,18333	0,23976
		Bias	-0,00026	0,00026	0,00018	-0,00017	0,00039	-0,00045	0,00008
		Errore St.	0,00414	0,00402	0,00419	0,00504	0,00433	0,00405	0,00486

Tab. 5.24: Relative Bias Modello Formativo-Formativo

Approccio	Path	100	250	500	1000
Ripetuto	β_1	-0,00267	-0,00221	-0,00215	-0,00118
	β_2	-0,00274	-0,00225	-0,00220	-0,00119
	β_3	-0,00267	-0,00238	-0,00222	-0,00119
	β_4	-0,00267	-0,00228	-0,00220	-0,00121
	β_5	-0,00278	-0,00234	-0,00222	-0,00120
	β_6	-0,00276	-0,00230	-0,00216	-0,00119
	β_7	-0,00274	-0,00221	-0,00222	-0,00117
Misto	β_1	-0,00259	-0,00199	-0,00229	-0,00124
	β_2	-0,00284	-0,00231	-0,00224	-0,00126
	β_3	-0,00268	-0,00238	-0,00213	-0,00114
	β_4	-0,00241	-0,00233	-0,00207	-0,00131
	β_5	-0,00252	-0,00229	-0,00220	-0,00110
	β_6	-0,00282	-0,00229	-0,00224	-0,00119
	β_7	-0,00307	-0,00236	-0,00216	-0,00105
PLS-R	β_1	0,02656	-0,00150	0,01585	-0,00648
	β_2	0,00669	0,00538	-0,00173	-0,00522
	β_3	-0,01660	-0,01031	-0,00313	-0,00457
	β_4	0,04385	0,00145	-0,00650	-0,00404
	β_5	-0,04756	-0,00021	-0,01587	-0,00052
	β_6	-0,03281	0,00083	-0,00846	0,00332
	β_7	-0,00520	-0,01274	0,00373	0,00708

5.4.3 Relative Bias

Nella Tabella 5.24 troviamo i valori degli RB per i *path* del modello Formativo-Formativo.

In questo caso gli RB per tutti i *path* e tutte le numerosità, pur diminuendo all'aumentare del campione e rimanendo sempre molto vicini allo zero, per l'approccio PLS-R risultano essere maggiori rispetto agli altri approcci. Nella Tabella 5.25 troviamo i valori degli RB per i *path* del modello Riflessivo-Riflessivo.

In questo caso, al contrario dello schema relazionale precedente (Formativo-Formativo in tabella 5.24) tutti i *path* (a parte eventi sporadici come nel caso del parametro β_5) vengono stimati meglio con l'approccio PLS-R. Passando agli schemi relazionali misti, nella Tabella 5.26 troviamo i valori degli RB per i *path* del modello Riflessivo-Formativo.

Anche in questo caso, la maggior parte dei β risulta essere stimata me-

Tab. 5.25: Relative Bias Modello Riflessivo-Riflessivo					
Approccio	Path	100	250	500	1000
Ripetuto	β_1	0,00297	0,00004	0,00169	0,00060
	β_2	0,00392	0,00107	0,00057	0,00098
	β_3	-0,01462	0,00064	0,00350	0,00030
	β_4	0,00287	0,00044	0,00241	0,00097
	β_5	0,00163	-0,00105	0,00206	-0,00032
	β_6	-0,00556	-0,00026	-0,00071	0,00082
	β_7	0,00063	0,00021	0,00123	0,00067
Misto	β_1	0,00251	0,00003	0,00176	0,00062
	β_2	0,00209	-0,00009	0,00023	0,00127
	β_3	-0,01479	0,00077	0,00321	0,00029
	β_4	0,00376	0,00072	0,00221	0,00055
	β_5	0,00175	-0,00096	0,00206	-0,00026
	β_6	-0,00535	0,00021	-0,00026	0,00117
	β_7	0,00172	0,00059	0,00186	0,00041
PLS-R	β_1	0,00226	0,00012	0,00213	0,00046
	β_2	0,00451	0,00135	0,00071	0,00080
	β_3	-0,01388	0,00075	0,00319	0,00026
	β_4	0,00319	0,00090	0,00172	0,00114
	β_5	0,00085	-0,00152	0,00233	-0,00032
	β_6	-0,00490	-0,00041	-0,00051	0,00103
	β_7	0,00021	-0,00003	0,00108	0,00064

Tab. 5.26: Relative Bias Modello Riflessivo-Formativo

Approccio	Path	100	250	500	1000
Ripetuto	β_1	0,00479	0,00385	0,00455	0,00021
	β_2	0,00713	0,00899	0,00125	0,00168
	β_3	0,00373	0,00508	0,00175	0,00148
	β_4	0,00902	0,00355	0,00078	0,00108
	β_5	0,00021	0,00540	0,00134	0,00060
	β_6	0,00535	0,00226	0,00108	0,00126
	β_7	0,00989	0,00512	0,00232	0,00001
Misto	β_1	0,00477	0,00382	0,00458	0,00023
	β_2	0,00703	0,00905	0,00124	0,00166
	β_3	0,00379	0,00503	0,00173	0,00151
	β_4	0,00909	0,00358	0,00082	0,00110
	β_5	0,00035	0,00533	0,00135	0,00059
	β_6	0,00525	0,00240	0,00108	0,00127
	β_7	0,00989	0,00505	0,00231	-0,00001
PLS-R	β_1	0,00393	0,00479	0,00414	-0,00025
	β_2	0,00519	0,00587	0,00110	0,00152
	β_3	-0,00096	0,00577	0,00023	0,00124
	β_4	0,00790	0,00221	0,00180	0,00045
	β_5	-0,00119	0,00260	0,00129	0,00113
	β_6	0,00883	0,00303	-0,00018	0,00119
	β_7	0,00940	0,00543	0,00259	0,00014

Tab. 5.27: Relative Bias Modello Formativo-Riflessivo

Approccio	Path	100	250	500	1000
Ripetuto	β_1	0,00479	0,00385	0,00455	0,00021
	β_2	0,00713	0,00899	0,00125	0,00168
	β_3	0,00373	0,00508	0,00175	0,00148
	β_4	0,00902	0,00355	0,00078	0,00108
	β_5	0,00021	0,00540	0,00134	0,00060
	β_6	0,00535	0,00226	0,00108	0,00126
	β_7	0,00989	0,00512	0,00232	0,00000
Misto	β_1	0,00477	0,00382	0,00458	0,00023
	β_2	0,00703	0,00905	0,00124	0,00166
	β_3	0,00379	0,00503	0,00173	0,00151
	β_4	0,00909	0,00358	0,00082	0,00110
	β_5	0,00035	0,00533	0,00135	0,00059
	β_6	0,00525	0,00240	0,00108	0,00127
	β_7	0,00989	0,00505	0,00231	-0,00001
PLS-R	β_1	0,00393	0,00479	0,00414	-0,00025
	β_2	0,00519	0,00587	0,00110	0,00152
	β_3	-0,00096	0,00577	0,00023	0,00124
	β_4	0,00790	0,00221	0,00180	0,00045
	β_5	-0,00119	0,00260	0,00129	0,00113
	β_6	0,00883	0,00303	-0,00018	0,00119
	β_7	0,00940	0,00543	0,00259	0,00014

glio con l'approccio PLS-R. Nella Tabella 5.27 troviamo i valori degli RB per i *path* del modello Formativo-Riflessivo.

Ancora una volta i β risultano essere stimati meglio con l'approccio PLS-R.

5.5 Conclusioni

Come già emerso, dopo aver presentato i vari approcci al PLS-PM di alto ordine, sia già noti in letteratura sia i nuovi sviluppi proposti, si è approfondito l'adattamento di tali approcci a tutti i tipi di relazioni tra variabili latenti e variabili manifeste, sia nel modello strutturale sia nel modello di misura. In questo capitolo si sono dunque effettuate varie simulazioni per valutare e confrontare i nuovi approcci proposti. I nuovi metodi proposti comparati all'approccio ripetuto hanno mostrato quasi sempre stime più stabili, sia al mutare della dimensione campionaria, sia al mutare delle relazioni tra le VL e le VM, sia al mutare del numero di VL del modello e delle VM per blocco. In tutti i casi presi in considerazione, la comunaltà è sempre più alta per l'approccio PLS-R. L'approccio PLS-R ha mostrato una variabilità più alta rispetto all'approccio misto per campioni piccoli. All'aumentare della numerosità campionaria e all'aumentare delle VL del modello, il PLS-R migliora in termini di stime, divenendo più stabile rispetto all'approccio misto. In termini di errore relativo, i risultati variano rispetto al numero di VL e VM del modello e rispetto al tipo di relazione che le lega.

Per quanto riguarda il modello con 3 VL, per lo schema Formativo-Formativo, il PLS-R ha delle stime peggiori, anche se vicine a quelle degli altri modelli e con numerosità molto ampia ($n=1000$) risulta addirittura migliore degli altri approcci. Sempre nel modello a 3 VL anche con schema Formativo-Riflessivo e Riflessivo-Formativo la situazione è simile alla precedente, con un miglioramento della variabilità dell'approccio PLS-R rispetto agli altri, all'aumentare del campione. Mentre nel modello Riflessivo-Riflessivo la situazione si inverte e il PLS-R risulta avere stime migliori degli altri approcci, tranne che per il campione da 1000 osservazioni.

Per quanto riguarda il modello con 4 VL, per lo schema Formativo-Formativo il PLS-R ha delle stime migliori degli altri modelli con qualsiasi numerosità campionaria. Con schema Riflessivo-Riflessivo la situazione è molto simile, il PLS-R risulta avere stime migliori degli altri approcci, ma in questo caso le stime sono estremamente vicine tra loro per i vari approcci. Sempre nel modello a 4 VL utilizzando invece uno schema Formativo-Riflessivo e Riflessivo-Formativo la situazione si inverte, con un peggioramento delle stime per l'approccio PLS-R rispetto agli altri, ma che tendono ad avvicinarsi all'aumentare del campione.

Infine, nel modello con 8 VL, la situazione è simile a quella per il modello a 4 VL, infatti per gli schemi Riflessivo-Riflessivo e Formativo-Formativo, gli errori di stima sono minori utilizzando il PLS-R, mentre per gli schemi

Riflessivo-Formativo e Formativo-Riflessivo, il modello misto ottiene stime più stabili.

In conclusione, i nuovi modelli proposti producono risultati abbastanza stabili in ogni situazione proposta, in più il PLS-R presenta sempre una comunaltà molto più alta degli altri approcci e ciò assicura un modello di misura meglio stimato.

I nuovi metodi sono sempre la scelta migliore, in termini di polarizzazione e MSE delle stime, quando il ricercatore mira a studiare le relazioni del modello con relazioni formative di costrutti del primo ordine al secondo ordine.

6. CUSTOMER SATISFACTION

6.1 *Introduzione*

La soddisfazione dei clienti è un asset fondamentale per ogni impresa. Un'accurata conoscenza del loro grado di soddisfazione è un aspetto altrettanto essenziale per perseguire il miglioramento continuo. Per tale motivo, da anni, molti soggetti sia pubblici che privati effettuano indagini di Customer Satisfaction, come strumento per interpretare il presente ed orientare il futuro.

La Customer Satisfaction è la soddisfazione del cliente in riferimento al bene ricevuto o al servizio erogato. Nelle organizzazioni che operano in regime di concorrenza e di mercato il fine che spinge a sviluppare un sempre maggiore orientamento al consumatore e alla valutazione della Customer Satisfaction è la ricerca della competitività. L'impresa ha bisogno di un consumatore soddisfatto, quindi di un cliente fedele [179]. Le ricerche di marketing dimostrano che è molto più difficile e costoso per le imprese riuscire ad ottenere nuovi clienti, piuttosto che trattenere quelli già presenti. Per questo aumentare la soddisfazione della clientela diventa un investimento redditizio e l'obiettivo da raggiungere per mantenere o incrementare quote di mercato. Le aziende pongono al centro della loro attenzione il cliente e cercano di gestire al meglio il rapporto stabilito con esso. Desiderano capire quali sono le sue esigenze ed aspettative e cercano di soddisfarle. L'esigenza che giustifica e rende necessario lo sviluppo di indagini di Customer Satisfaction è quella di ascoltare e comprendere a fondo i bisogni che il cliente esprime, porre attenzione costante al suo giudizio, sviluppare e migliorare la capacità di dialogo e di relazione. Tuttavia, poichè ci sono diverse tipologie di clienti, risulta difficile soddisfare i desideri di ciascuna di queste. Per questo motivo bisogna capire a fondo le difficoltà, le insoddisfazioni e i bisogni non colmati di ogni singolo cliente, c'è la necessità di comprendere il cliente inappagato e di riconoscere il tipo di lamentela. Nell'attuale realtà produttiva, dove i prodotti sono sostanzialmente tutti uguali e dove anche i prezzi di mercato tendono a ruotare intorno a valori medi, la soddisfazione del cliente e la sua fedeltà assumono fondamentale

importanza e non vengono più determinate esclusivamente dalla qualità del prodotto e dal rapporto prezzo-qualità. Il cliente non è attento solo alle caratteristiche del bene, ma anche ai servizi che l'azienda produttrice è in grado di fornirgli in relazione all'acquisto effettuato. Esso stesso non è più fruitore passivo, ma diventa committente di prestazioni in grado di co-determinare le scelte strategiche del management aziendale. La soddisfazione dei clienti deve essere oggetto di continue misurazioni e i risultati di tali misurazioni devono determinare le scelte utili al perseguimento degli obiettivi.

6.2 *Perché rilevare la Customer Satisfaction*

La realizzazione di un'indagine di Customer Satisfaction può essere finalizzata sia a definire e valutare l'efficacia delle "politiche aziendali" sia a progettare i sistemi di erogazione dei servizi o dei beni. Per politiche aziendali si intendono, in modo sintetico, le scelte prioritarie del management finalizzate ad ottenere benefici esterni sulla comunità. La realizzazione delle politiche aziendali comporta la definizione di:

- Direzione da seguire
- Obiettivi da raggiungere
- Vincoli entro cui operare
- Linee guida (come agire per ottenere gli obiettivi)
- Valutazione dell'impatto

La Customer Satisfaction può diventare, nella logica della qualità, uno strumento molto utile nella scelta delle priorità e nella verifica dell'efficacia delle azioni intraprese in quanto:

- Può rappresentare ed evidenziare i bisogni e le attese degli utenti
I bisogni e le attese degli utenti, attraverso la Customer Satisfaction, possono essere ben individuati e focalizzati, ognuno di essi è ordinato per priorità ed è definito in termini di prestazione minima accettabile e di prestazione ideale. Inoltre la percezione viene confrontata con indicazioni relative alle attese, il che consente di individuare dove concentrare gli sforzi. Le priorità vengono definite dai bisogni ritenuti importanti dai cittadini per i quali il livello di soddisfazione non è in linea con le attese.

- Può favorire la comprensione dei bisogni latenti
Questa tendenza verso la soddisfazione degli utenti può sviluppare nelle imprese la sensibilità e le capacità di cogliere i segnali deboli, di anticipare i bisogni, di scoprire i bisogni latenti. La capacità di comprensione dei bisogni latenti costituisce un forte stimolo all'innovazione dei servizi e alla definizione di nuove risposte ai bisogni.
- Può aiutare a cogliere idee, spunti e suggerimenti
L'ascolto attento è una fonte inesauribile di proposte, suggerimenti, stimoli alla definizione di interventi sempre più efficaci. La Customer Satisfaction può coniugare il flusso di informazioni che proviene dall'esterno con quello che proviene dall'interno.
- Può facilitare il superamento dei vincoli interni
A volte i vincoli interni dati dall'agire ripetitivo e ciclico sono un freno, invisibile ma potente, alla capacità delle amministrazioni di orientare l'azione verso l'utente. La Customer Satisfaction, portando in maniera oggettiva il punto di vista dell'utente all'interno dell'organizzazione, facilita scelte che sappiano andare oltre questi vincoli.
- Può supportare la verifica e la comprensione dell'efficacia delle politiche
Il monitoraggio sistematico, nel tempo, del livello di soddisfazione rende possibile la verifica di come l'azienda possa riuscire ad affinare le capacità di rispondere agli utenti. Le evoluzioni delle percezioni consentono di capire in quale misura le politiche attivate hanno raggiunto i loro scopi ed hanno inciso veramente all'esterno.

Possiamo quindi affermare che la Customer Satisfaction consente di migliorare l'erogazione dei beni/servizi guidati dai bisogni degli utenti, utilizzando al meglio le risorse disponibili.

6.3 *Come rilevare la Customer Satisfaction*

Le fasi di una rilevazione di Customer Satisfaction, in termini analitici possono essere così descritte:

1. La preparazione della rilevazione
In questa fase si deve, innanzitutto, definire l'obiettivo dell'indagine e si dovrà decidere se realizzare la ricerca all'interno dell'organizzazione, oppure rivolgersi ad un consulente/agenzia esterna. A questo

punto si procederà alla realizzazione di una pre-indagine esplorativa scegliendo il metodo di rilevazione più adeguato all'obiettivo.

2. La raccolta dei dati

La raccolta dei dati sarà anticipata da una possibile fase qualitativa (interviste, focus group, ecc.) attraverso la quale si definiranno meglio i confini e le caratteristiche delle problematiche oggetto dell'indagine. Al termine di questa prima, possibile, fase di avvio, si procederà con la ricerca quantitativa scegliendo le modalità di somministrazione, creando l'intervista, definendo l'eventuale campione e procedendo all'esecuzione delle interviste.

3. L'elaborazione e l'interpretazione dei dati

E' la fase nella quale si elaborano i dati raccolti per comprendere se gli utenti sono soddisfatti o non sono soddisfatti dei servizi erogati. In questa fase è possibile creare un ordine di priorità alla soddisfazione/insoddisfazione e individuarne le cause principali definendo delle strategie di miglioramento.

4. La presentazione e l'utilizzo dei risultati

E' la fase conclusiva, nella quale si procede alla presentazione dei risultati e si avviano attività e progetti di miglioramento in funzione di una maggiore soddisfazione degli utenti.

La selezione dei servizi oggetto d'indagine costituisce la prima decisione dell'organizzazione e la prima fase del processo per rilevare la soddisfazione e costruire i relativi indicatori di prestazione. Un'indagine completa rileva la soddisfazione e misura i relativi indicatori di prestazione su ciascun servizio e su tutti gli utenti. L'indagine può essere altresì limitata ad un campo definito, per esempio un segmento di clientela o un solo servizio o una specifica area dell'organizzazione. Uno dei momenti più critici, per il successo di un'indagine di Customer Satisfaction, è la fase iniziale e soprattutto il momento in cui si stabiliscono gli obiettivi, se possibile in modo definitivo. È di fondamentale importanza avere le idee molto chiare sugli obiettivi dell'indagine e interrogarsi a fondo per identificare, nei minimi dettagli, i traguardi conoscitivi a cui si vuole giungere. L'individuazione di questi obiettivi deve essere oculata e precisa: ogni obiettivo si traduce in un costo e ciò vale sia nel caso che l'indagine venga svolta all'interno dell'ente erogatore del servizio, sia nel caso in cui intervenga un'organizzazione esterna. E' da tener in considerazione anche la scelta del modello da applicare per la rilevazione dei dati. Nel corso degli anni, infatti, si sono

sviluppate diverse metodologie per l'analisi della soddisfazione del cliente, in sostituzione di quelle più tradizionali, oggi considerate piuttosto approssimative, riferite alla gestione e al conteggio dei reclami, oppure, nelle organizzazioni profit, alla fedeltà del cliente [179]. Nel capitolo successivo si esporranno alcuni modelli operativi ritenuti particolarmente adatti alla valutazione della Customer Satisfaction nell'ambito dei servizi. Dopo aver definito gli obiettivi, gli ambiti di intervento, le segmentazioni degli utenti ed il modello di riferimento, si procede nella rilevazione, che avviene tramite questionario.

6.4 *Come costruire il questionario*

Nel corso della pianificazione di un'indagine, una volta che sia stata individuata la popolazione oggetto di studio e si sia decisa la tecnica di rilevazione, si realizza una delle fasi più delicate: la costruzione del questionario. Essa richiede una efficace progettazione, in quanto il questionario è uno schema di una intervista altamente strutturata, il cui scopo è quello di raccogliere informazioni sulle variabili, qualitative e quantitative, oggetto dell'indagine. Il questionario è uno strumento di rilevazione molto adoperato nelle indagini statistiche; può essere considerato sia come strumento di misura che come strumento di comunicazione, infatti esso deve anche trasmettere all'intervistato il significato esatto dell'informazione richiesta. Affinchè un questionario possa garantire la confrontabilità delle informazioni ricevute si devono imporre due condizioni alle domande, esse devono essere rivolte a tutti nelle stesse condizioni e devono avere lo stesso significato per tutti i fruitori del questionario[13]. Le fasi per la costruzione di un questionario sono tre:

1. Preparazione dello schema concettuale
2. Formulazione
3. Verifica

La preparazione dello schema concettuale del questionario, che avviene nella prima fase, consiste nella conversione delle esigenze conoscitive, espresse genericamente, nei concreti operativi dell'indagine, ossia nelle domande relative ai sottoproblemi che via via emergono dal problema generale. Si cerca di ricondurre il fenomeno a cui si è interessati ad un modello logico-concettuale. La struttura del questionario viene progettata anche in relazione alla tecnica utilizzata per l'intervista. Questa può essere:

- Intervista diretta
- Intervista telefonica
- Autocompilazione

La seconda fase consiste nella formulazione delle domande di diversa tipologia: quelle specifiche dell'indagine o quelle utilizzate per i controlli di qualità sui codici identificativi ed infine quelle sulle caratteristiche strutturali degli individui. Un altro punto importante nella progettazione del questionario è la sequenza delle domande per una stessa tematica. Le domande possono essere distinte:

- A) per come sono formulate:
abbiamo quindi, domande dirette, quando si chiama in causa direttamente l'intervistato, e domande indirette, quando ci si riferisce ad un generica terza persona o ad un gruppo di persone con caratteristiche simili.
- B) per il tipo di risposte previste:
possiamo distinguere domande a risposta semplice, che ammettono solo una risposta tra quelle possibili, e domande a risposta multipla, che consentono di scegliere più di una risposta anche se incompatibili tra di loro.

Esistono anche domande aperte nelle quali si lascia all'intervistato la piena libertà di risposta e, in contrapposizione, quelle chiuse, dove si considera un numero limitato di alternative predefinite di risposta tra le quali l'intervistato deve scegliere quella che si avvicina il più possibile al suo pensiero. Le domande di un questionario devono avere tre importanti qualità:

1. devono focalizzare il più possibile l'argomento
2. devono essere brevi
3. devono essere chiare

La terza fase ha una duplice funzione: da un lato verifica la contiguità tra lo strumento di misurazione adottato nella progettazione del questionario e le esigenze conoscitive dell'indagine, dall'altro valuta la funzionalità del questionario sia come strumento di comunicazione che come strumento per l'intervistatore. Durante la fase della predisposizione dell'indagine e della rilevazione si può incorrere in vari tipi di errori non campionari, ossia non dovuti al campionamento:

1. Errori dovuti al ricercatore (ad esempio la non chiarezza degli obiettivi dell'indagine e dell'unità di rilevazione o una erronea formulazione del questionario o un'eccessiva lunghezza di esso o della sequenza delle domande)
2. Errori dovuti a coloro che rispondono (ad esempio la mancanza di informazione, scarso interesse a rispondere o mancanza di memoria o la non conoscenza degli argomenti trattati nell'intervista)
3. Errori dovuti all'intervistatore (ad esempio incompetenza nella conduzione dell'intervista, o impreparazione sul questionario, oppure il condizionamento alle risposte con la sua presenza)
4. Errori dovuti all'operazione di registrazione (ad esempio la registrazione sbagliata del valore del dato)

Alcuni di questi errori non dipendono dal ricercatore, altri invece sì. In particolare una formulazione adeguata del questionario è fondamentale, in quanto tale strumento consente di trasformare i dati rilevati dalle unità statistiche in informazioni da utilizzare per le elaborazioni successive. Bisogna evitare di formulare domande troppo generiche o troppo tecniche e quindi comprensibili solo da una parte limitata di persone. La successione delle domande non deve influenzare le risposte alle domande successive. Le domande non devono essere formulate in modo da indirizzare l'intervistato verso particolari risposte, infine non devono creare stati di imbarazzo tra l'intervistato e l'intervistatore. Il questionario, quindi può diventare uno strumento fondamentale per la prevenzione di errori e per la loro limitazione, sia per quelli dovuti all'intervistatore sia per quelli causati da coloro che rispondono sia per quelli dovuti all'operazione di registrazione.

6.4.1 *Le scale di misurazione*

Nella costruzione del questionario è anche importante prestare attenzione alle scale di misurazione degli atteggiamenti e delle opinioni. Le scale si possono dividere in vari tipi, abbiamo quindi:

1. SCALE NOMINALI (Contraddizione in termini)
2. SCALE ORDINALI (Maggiore/uguale/minore)
3. SCALE AD INTERVALLI (Mancanza dello zero assoluto)
4. SCALE CARDINALI (Numeri reali: discreti e continui)

1. Scale nominali:

la denominazione “nominali” costituisce una contraddizione in termini, in quanto le proprietà nominali (categoriali, di carattere qualitativo) non sono collocabili lungo una scala. Ad esempio, proprietà come la professione, la religione, le opinioni, ecc. non sono collocabili lungo una scala perché ogni posizione non è “più” o dopo o prima di un’altra. Essendo delle proprietà di tipo qualitativo, infatti, non possiamo dire, ad esempio, che “cattolico” è “più” di “protestante” e “meno” di “musulmano”. Ogni stato sulla proprietà ha una sua autonomia semantica e una sua indipendenza. Dunque, trattandosi di una scala che si basa sulla registrazione di dati di carattere qualitativo, è più opportuno utilizzare il termine scala tra virgolette. I dati solo in alcuni casi possono essere trasformati in variabili quantitative. Ciò accade, ad esempio, quando è possibile trasformare una variabile categoriale in dicotomica, come nel caso del genere, dove è possibile assegnare un valore numerico a M/F (ad esempio: M=1; F=0); in questo caso la variabile è stata trasformata in una di tipo metrico che ci permette il conteggio del numero di stati sulla proprietà.

2. Scale ordinali (o con categorie ordinate):

Le proprietà categoriali ordinate o non hanno autonomia semantica o comunque hanno un’autonomia semantica ridotta rispetto alle categoriali non ordinate. Secondo Campbell [18] si ha un livello ordinale quando, in assenza di un’unità di misura, l’ammontare della proprietà x posseduta dall’oggetto A è confrontata con l’ammontare della proprietà x posseduta dall’oggetto B e l’esito del confronto è un giudizio di maggiore/uguale/minore. Un esempio di scala con categorie ordinate è la Likert (la prima versione formulata nel 1934 era a 7 intervalli; oggi, quella più in uso nella ricerca sociologica, è a 5 intervalli):

Rispetto alle categoriali non ordinate disponiamo, certamente, di più informazioni. Ogni categoria, infatti, ha una sua caratteristica ed è equidistante dalle altre. L’informazione in più di cui disponiamo con la scala ordinale è rappresentabile, invece, lungo una vera e propria scala:

Quello che non ci è consentito sapere è la distanza tra una posizione e un’altra. In qualche modo, per alcuni tipi di variabili, la scala ad intervalli risolve questo tipo di problema. E’ possibile cioè assegnare dei valori numerici alle distanze, attraverso l’uso di tecniche piuttosto complesse. Tra i maggiori vantaggi dello strumento abbiamo la facilità di codifica e di decodifica delle informazioni. Ad esempio: 7; 5; 0; 3; 1; costituiscono dei codici o etichette numeriche facilmente utilizzabili, che non hanno però proprietà cardinali. Questo livello di scala, quindi, ci permette oltre che di stabilire delle differenze, di definire delle relazioni asimmetriche d’ordine anche se

non conosciamo la larghezza degli intervalli.

3. Scale quasi-cardinali (o ad intervalli):

le proprietà quasi cardinali sono proprietà per le quali non esiste un'unità di misura universalmente accettata. E' però possibile, con alcuni accorgimenti (attraverso l'uso di tecniche note come tecniche di scaling), ricondurle a proprietà di tipo cardinale e dunque utilizzare le stesse operazioni utilizzate a livello della scala a rapporto. Proprietà di tipo qualitativo, quindi, vengono trasformate in proprietà di tipo quantitativo. Ad esempio: la temperatura è, di per sé, una variabile che ha caratteristiche di tipo continuo. Esistono diversi modi di misurare la temperatura tra i quali l'uso di una scala che utilizza lo zero assoluto. In questo caso, siamo di fronte ad una proprietà propriamente cardinale. La scala più utilizzata nel nostro paese è, però, la Celsius dove lo zero non è assoluto ma relativo, in quanto convenzionalmente stabilito (coincide con la temperatura di congelamento dell'acqua). In quest'ultimo caso non è più possibile parlare di scala a rapporto, ma è necessario parlare di scala ad intervallo perché la caratteristica di suddetta scala è proprio quella della mancanza di uno zero assoluto anche se è definita la grandezza dell'intervallo che separa un valore dagli altri in quanto esiste un'unità di misura che vale a tutti i livelli della scala. Avendo uno zero arbitrario, come in tutte le misurazioni in scale di intervallo, possiamo operare sui valori utilizzando molte delle operazioni aritmetiche che sono disponibili; ma certo non possiamo fare rapporti. Infatti, affinché l'operazione di rapporto abbia senso, occorre che l'origine della scala sia fissa e non arbitraria. Un esempio di questo tipo di scala è la Thurstone che è una scala ad 11 punti, dove, fissati i punti estremi, il punto medio è stabilito che gli intervalli sono tra loro equivalenti, esperti vengono chiamati a fare delle valutazioni separatamente sugli items proposti. Ogni esperto, cioè, deve mettere insieme 11 gruppi di affermazioni per somiglianza e distribuirli lungo la scala. Lo sperimentatore, eliminate le affermazioni su cui non vi è convergenza di scelta, costruirà il questionario che utilizzerà nella ricerca.

4. Scale cardinali (o a rapporto):

per proprietà cardinali intendiamo delle proprietà che hanno un cardine rappresentato dallo zero, a destra e a sinistra del quale troviamo valori positivi e valori negativi. Si suddividono in discrete e continue. Le proprietà cardinali discrete sono quelle che enumeriamo o conteggiamo. Non sono misurabili in quanto non si collocano lungo un continuum. Ciascuno di noi, ad esempio, costituisce una proprietà di tipo discreto. Non esiste una misurazione del numero di individui, ma solo un'enumerazione. Quello che si fa in questi casi (come, ad esempio, nell'ambito della statistica de-

scrittiva) è contare: 1, 2, 3, 4... (con il solo utilizzo dei numeri naturali). Cosa diversa costituiscono, invece, proprietà come la lunghezza, il peso, la densità, ecc. che costituiscono le cosiddette proprietà continue. Proprietà, cioè, collocabili lungo un continuum e che possiamo misurare. L'atto di misurare, poiché il potere risolutivo dei nostri strumenti di misurazione è comunque limitato, le rende in qualche modo discrete perché non è possibile scrivere un numero con infinite cifre decimali come una proprietà continua. Dunque, è inevitabile una certa approssimazione ed è il motivo per cui esistono delle regole di arrotondamento. Ad esempio nel dire l'età non indichiamo mese, giorno, ora, minuto, ecc. ma ci limitiamo all'anno che non è altro che una forma di arrotondamento. A questo livello di scala, poiché lo zero non è arbitrario ma assoluto (coincide, cioè, con l'assenza della proprietà in questione), sono possibili tutte le operazioni definite sui numeri reali [42].

6.5 *Principali limiti del Questionario*

L'analisi della customer satisfaction ha sempre presentato molte problematiche, dalla scelta del questionario alla scelta del metodo di somministrazione dello stesso. Abbiamo già parlato di come contruire un questionario ma anche la scelta del metodo di somministrazione è altrettanto importante.

Di seguito analizzeremo brevemente le tecniche principali di somministrazione, quali le interviste faccia a faccia, le interviste telefoniche, i questionari postali o tramite fax/mail e i questionari on-line. Ogni soluzione presenta dei vantaggi e degli svantaggi rispetto alle altre. La scelta della tecnica di raccolta più idonea è molto legata agli obiettivi dell'indagine, alle caratteristiche dei soggetti di analisi e, ovviamente, ai vincoli di tempi e costi.

- **Interviste faccia a faccia (face to face)**

Le interviste faccia a faccia sono indicate nei casi in cui occorre affrontare argomenti complicati, lunghi o che richiedano di far vedere o far testare qualcosa, o quando i clienti sono molto concentrati a livello geografico. Nelle interviste personali l'intervistatore stabilisce spesso un rapporto con la persona intervistata, ciò crea dell'empatia tra intervistato ed intervistatore e questo può aumentare l'interesse ed evitare equivoci legati ad errate comprensioni delle domande. Lo svantaggio principale delle interviste faccia a faccia è che potrebbe verificarsi un'influenza sulle risposte da parte dell'intervistatore; inoltre, questo

metodo di rilevazione comporta alti costi e tempi maggiori rispetto agli altri metodi.

- Interviste telefoniche

Le interviste telefoniche sono utilizzate efficacemente in moltissimi settori ed offrono il vantaggio di svolgersi rapidamente e da un luogo univoco; rappresentano, infatti, il metodo generalmente utilizzato per sondare campioni numerosi. I vantaggi di questa metodologia di rilevazione sono i costi medio-bassi di somministrazione, l'ottimo controllo delle caratteristiche del campione e brevi tempi richiesti per la realizzazione. Tuttavia, uno svantaggio importante delle indagini telefoniche è dato dalla difficoltà nel proporre troppe domande o troppo lunghe. In questi casi gli intervistati potrebbero essere portati ad interrompere la telefonata. Inoltre, alcune tipologie di clienti sono difficili da contattare telefonicamente; mentre, altri, si rifiutano di rispondere a causa dell'intenso utilizzo del telemarketing, che ha provocato fastidio all'utenza telefonica.

- Questionari postali o e-Mail

I questionari spediti tramite posta o e-mail possono essere utilizzati abbastanza efficacemente quando c'è un grande interesse dei clienti per l'oggetto dell'indagine. Questo metodo di rilevazione offre il vantaggio di avere bassi costi e semplicità di gestione, inoltre i clienti sono liberi di scegliere quando rispondere. Spesso tali tipi di questionari hanno una bassa percentuale di risposta, ed inoltre chi risponde, spesso è un cliente "estremo" quindi o molto soddisfatto o molto insoddisfatto; per questo il campione potrebbe non risultare rappresentativo della popolazione. I questionari postali o tramite fax negli ultimi anni sono sempre meno utilizzati, in quanto viene preferita la compilazione dei questionari via web.

- Questionari on-line

La somministrazione attraverso internet è un metodo che consente di ridurre i costi (viene eliminato il costo degli intervistatori ed il costo di imputazione dei dati, rispetto alle interviste face to face) ed offre il vantaggio di essere impersonale. Un ulteriore vantaggio è la scelta del momento di compilazione da parte dell'intervistato e non da parte dell'intervistatore. Tuttavia, le interviste via web presentano il problema della significatività del campione. Quindi, per dare una rappresentatività corretta rispetto all'universo di riferimento, è necessario ricreare un campione adeguato prima dell'analisi, con il rischio

di perdere informazioni. Inoltre, anche questo metodo di rilevazione, come quello postale, ha lo svantaggio di ottenere basse percentuali di risposta. La tendenza è comunque quella di privilegiare il questionario on-line sia perché la rilevazione risulta meno costosa sia perché tale metodo consente il monitoraggio continuo della situazione. Inoltre, con i questionari online è possibile arricchire il questionario con filmati o immagini che permettono di ottenere un questionario più chiaro e accattivante con bassi costi [165].

Tutte queste tipologie di somministrazione del questionario portano con sé il problema della lunghezza del questionario. Anche se in maniera diversa tutti queste tipologie di questionario gioverebbero della riduzione del numero di domande.

Questo tipo di problematica, infatti, porta con sé questioni sia relative al tempo di somministrazione, sia al relativo rischio di abbandono, sia problematiche di natura economica; si pensi, come già affermato, ai costi di un call center o degli intervistatori.

Tra i vari studiosi che si sono occupati di questa tematica molti hanno posto l'attenzione al tasso di abbandono dei rispondenti (percentuale di individui che abbandonano l'intervista). Ad esempio MacElroy [110], spiega quali siano i quattro principali fattori che influenzano il tasso di abbandono:

- Lunghezza del questionario, ovvero il numero di schermate necessarie a completare il questionario (una pagina video può contenere una o più domande)
- Tempo medio per completare un questionario
- Il valore totale degli incentivi ai rispondenti;
- Il valore degli incentivi indicati nella lettera di invito o nella pagina di introduzione al questionario

Poc'hè incentivare i clienti ad una risposta è prassi ormai poco utilizzata, soprattutto quando si tratta di clienti *consumer*, ci soffermeremo su i primi due fattori. In tabella 6.2 si può notare come, secondo lo studio di MacElroy, la diminuzione del tasso di abbandono è notevole soprattutto nel passare da un tempo di compilazione di 15 minuti a 10.

In questa direzione va anche lo studio di Hoerger [76] in cui utilizzando tecniche di analisi della sopravvivenza si arriva a definire una formula di percentuale perduta di abbandono:

Tab. 6.1: Tasso di abbandono per numero di pagine

Numero di schermate	Tasso di Abbandono
45	73%
40	56%
35	42%
30	30%
25	20%
20	13%
15	9%
10	7%

Tab. 6.2: Tasso di abbandono per minuti di compilazione

Tempo medio di compilazione	Tasso di Abbandono
30	50%
25	44%
20	35%
15	24%
10	9%

$$PPD = 10\% + 2\% * (100 - items) \quad (6.1)$$

Tale formula è stata pensata per calcolare a priori il numero di partecipanti da contattare in modo da raggiungere il target del campione, ma dimostra al tempo stesso che all'aumentare del numero di domande il tasso di abbandono cresce sensibilmente. Questi sono solo alcuni esempi di quanto la riduzione della lunghezza del questionario può essere importante per ottenere un cospicuo campione di rispondenti.

Per rispondere a questi problemi, ad oggi, molti dei soggetti che si occupano di customer satisfaction utilizzano il *Net Promoter Score*, che consiste in un indice ottenuto dall'analisi dei risultati di un'unica domanda, che va a sostituire dunque l'intero questionario.

6.6 NPS - Net Promoter Score

L'NPS (*Net Promoter Score*) è un indice che può essere utilizzato per valutare la fedeltà del cliente ad un Prodotto/Azienda. In alcuni casi è una

valida alternativa alla tradizionale analisi di soddisfazione del cliente. Questo strumento è una metrica sviluppata da Fred Reichheld [140], introdotto nel 2003. L' NPS è calcolato in base alle risposte ad una singola domanda: "Quanto consiglieresti la nostra compagnia/prodotto/servizio ad un amico o un collega?" La risposta a questa domanda è basata su una scala da 0 a 10. I Clienti che rispondono con un valore tra 9 e 10 vengono considerati "promoters", quelli che rispondono con valori tra 7 e 8 vengono considerati "passive" e gli altri vengono considerati "detractors". Per calcolare l'indice si calcolano prima la percentuale di "promoters" sul totale, poi la percentuale di "detractors".

L'Nps è definito con la seguente formula:

$$\text{PercentualePromoters} - \text{PercentualeDetractors} \quad (6.2)$$

L'NPS dunque può assumere valori da -100 (tutti i clienti sono "detractors") a +100 (tutti i clienti sono "promoters"). Un valore +50, viene considerato un valore eccellente.

6.7 Conclusioni

Come visto, l'analisi NPS permette di avere un indice facile da elaborare che riduce tempi e costi della raccolta dati. Questo tipo di indicatore risulta quindi utilissimo per tenere sotto controllo l'andamento della soddisfazione dei soggetti di analisi. Si tratta però di un indicatore di performance, e dunque, quando questo indicatore risulta basso, diventa impossibile capire quali siano gli aspetti che rendono insoddisfatto il cliente. In conclusione, in questi casi, l'analisi della customer satisfaction, pur essendo molto laboriosa e dispendiosa, resta uno dei pochi strumenti validi per comprendere a fondo i desideri dei clienti, rispondere alle loro esigenze e valorizzare le loro esperienze. Di conseguenza si rende necessario affrontare le problematiche riguardanti la Customer Satisfaction in maniera diversa. Come fare allora in questi casi ad accorciare il questionario ?

Non volendo eliminare le domande dirette agli aspetti tangibili del prodotto / servizio / azienda si propone di eliminare le domande generiche riguardanti la soddisfazione del cliente. Considerando gli studi sui modelli gerarchici di alto ordine, si può riportare il modello di CS ad uno di questi modelli, considerando appunto la soddisfazione come una variabile latente di cui non è possibile misurare alcuna variabile manifesta. Ciò permette

di accorciare il questionario eliminando tra le 3 e le 6 domande. Nel capitolo successivo si illustrerà un caso applicativo sia dell'utilizzo dell'NPS sia dell'utilizzo di un modello gerarchico, confrontando i risultati ottenuti dall'applicazione di tali modelli, con un'analisi eseguita utilizzando un algoritmo di PLS-PM classico.

7. CASO APPLICATIVO SU UN'AZIENDA MULTYUTILITY

Come definito da Bruti Liberati [14] “con l’espressione multiutility ci si riferisce alla struttura che le imprese di servizi di pubblica utilità, siano esse pubbliche o private, assumono al termine di un graduale processo di convergenza, che le porta ad operare contemporaneamente in più settori (energia elettrica, gas naturale, acqua, telecomunicazioni) e con una forte presenza nelle fasi finali di distribuzione e vendita. Alimentato in gran parte dai processi di liberalizzazione e privatizzazione dei servizi pubblici e dal conseguente processo di riorganizzazione aziendale che ha riguardato tutti i principali operatori del settore, il fenomeno delle multiutility ha indubbiamente ricevuto una forte accelerazione dall’apertura dei mercati nazionali dell’energia. [...] All’intensificarsi della concorrenza indotta dalla progressiva *deregulation* di questi mercati, infatti, le imprese di pubblici servizi hanno risposto modificando la struttura del proprio business nel tentativo di limitare, da un lato, il calo dei margini di profitto precedentemente realizzati in condizioni di monopolio e, dall’altro, di sfruttare attività con rendimenti più alti.” Ad ogni modo, al centro del modello multiutility vi è la consapevolezza che per una società di pubblici servizi il rapporto con il cliente costituisce l’asset di maggior valore e che la combinazione tra una vasta base di clienti ed un marchio fortemente riconosciuto dal mercato risulta vincente, così come vi è la consapevolezza che i vantaggi strategici e di efficienza di questa scelta richiedono una forte focalizzazione nella fase finale di distribuzione, commercializzazione e vendita dei servizi. La crescente diffusione del modello dell’azienda multiservizio in Italia e l’imminente chiusura del mercato tutelato¹ hanno portato sul mercato una forte concorrenza e dunque diviene di fondamentale importanza per gli operatori del settore fidelizzare il cliente e tenerne sotto controllo la soddisfazione. In questo tipo di scenario, si è proposto ad una media azienda multiutility, operante su tutto il territorio italiano, di affrontare un’analisi di controllo della soddisfazione dei propri clienti. Il primo passo scelto dalla multiutility oggetto di analisi, è stato quello di monitorare la soddisfazione dei propri

¹ Al momento la chiusura del mercato tutelato è prevista per il Primo Gennaio 2019.

clienti tramite l'indice NPS.

7.1 Applicazione dell'NPS

Per rilevare e analizzare l'indice di Net Promoter Score, tenendo conto anche di questioni di budget e di scadenze, si è deciso inviare a tutti i clienti di cui si era in possesso di una mail o di un numero di telefono *mobile*, una mail e/o un sms contenenti il link ad uno strumento di rilevazione online, in cui i clienti potessero risondere alla domanda "Quanto consiglieresti la nostra azienda ad un amico o un collega?" con una valutazione da 0 a 10 (Si veda il paragrafo 6.6).

Trascorsa una settimana i dati sono stati analizzati e dopo aver ripulito la base dati da risposte di clienti inesistenti o da risposte plurime, sono state raccolte 526 interviste.

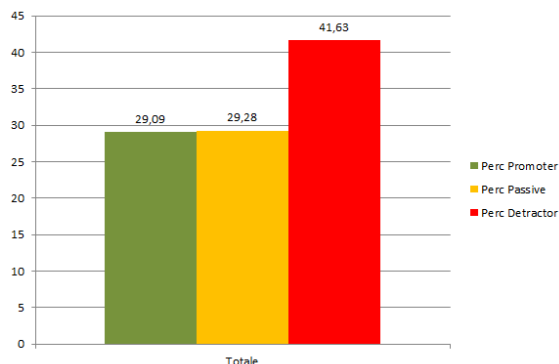


Fig. 7.1: Percentuali totali

Nella figura 7.1 possiamo vedere i primi risultati di analisi con le percentuali dei clienti divisi tra "Promoter", "Passive" e "Detractor". Già ad un primo sguardo si può notare come purtroppo la percentuale di "Detractor" sia più alta rispetto alle altre due.

Per ottenere ulteriori informazioni utili si è deciso di distinguere i clienti in *Business* e *Consumer*. In figura 7.2 si può notare come per i clienti *Consumer* la percentuale di "Detractor" sia più bassa rispetto ai clienti *Business*, restando però leggermente più alta di quelle dei "Promoter".

Si è infine deciso di analizzare anche la provenienza geografica dei clienti.

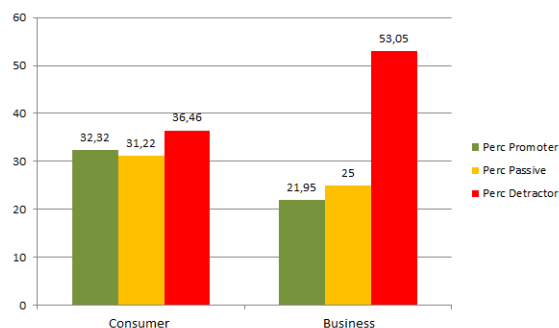


Fig. 7.2: Percentuali per tipo cliente

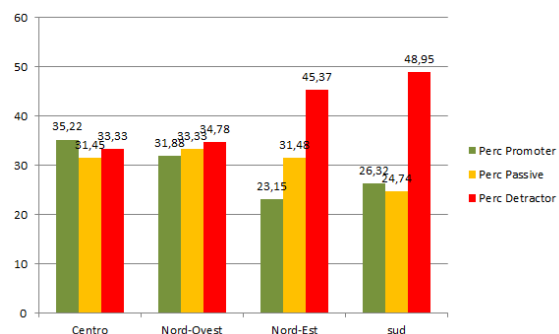


Fig. 7.3: Percentuali per area geografica

In questo caso, in figura 7.3 si può notare come al centro ci sia una leggera maggioranza di clienti “Promoter”, mentre nelle altre aree geografiche resti una preponderanza di clienti “Detractor”, soprattutto al sud Italia.

Anche se i risultati analizzati per cluster hanno dato spunti interessanti per analisi future, si è deciso di procedere all’analisi non suddividendola in gruppi. Calcolando quindi l’indice NPS sui dati raccolti si è ottenuto un indice di -12,54.

L’indice ottenuto, pur non essendo estremamente negativo, indica comunque una poca soddisfazione da parte del cliente. Per entrare dunque nel merito di questa poca soddisfazione si è deciso di effettuare anche una analisi di Customer Satisfaction classica.

7.2 Applicazione della Customer Satisfaction

Come detto nei paragrafi precedenti, la ricerca sulla soddisfazione, che di seguito affronteremo nelle sue peculiarità, ha come scopo quello di valutare gli “outcome” aziendali, tentando di comprendere le percezioni dei clienti in tema di qualità dei servizi loro offerti. Di conseguenza, il principale risultato che si è voluto raggiungere è stato quello di costruire e somministrare un questionario sulla soddisfazione che potesse indirizzare le strategie aziendali di miglioramento. Per poter rilevare la soddisfazione in maniera più rapida ed economica possibile, il questionario è stato proposto a tutti i clienti di cui si era in possesso di una mail o di un numero di telefono *mobile*. A questi clienti sono stati inviati un messaggio Sms e/o una mail contenenti il link al questionario, con una breve descrizione dello stesso. I dati sono stati raccolti durante il mese di Marzo 2017, trascorse due settimane i dati sono stati analizzati e dopo aver ripulito la base dati da risposte di clienti inesistenti o da questionari incompleti, sono state raccolte 600 interviste.

7.2.1 Il questionario

L'ipotesi di partenza è stata che i clienti formino il proprio giudizio sulla qualità del servizio valutando le stesse dimensioni sulle quali si costruisce il giudizio degli utenti in generale e aggiungano a queste, altre valutazioni inerenti le caratteristiche, le tappe e le criticità peculiari della propria esperienza “aziendale”. Il questionario è stato costruito sulla base di esperienze pregresse², coadiuvate da un'analisi della struttura aziendale e da un'ulteriore ricerca bibliografica su strumenti già esistenti di rilevazione della percezione della qualità dei servizi³. Le aree tematiche identificate sono state 5 ed hanno coperto sia le dimensioni ritenute fondamentali in letteratura sia le aree aziendali critiche.

² Tesi magistrale e triennale

³ Si veda tra gli altri: ECSI Technical Committee (1998), European Customer Satisfaction Index, Foundation and Structure for Harmonized National Pilot Projects. Report.; ESIS, (2001-2004), European Satisfaction Index System, IST Project within the Vth Framework Programme (IST-2000-31071) of the European Commission.

In particolare:

- Area 1: Il Consulente (5 item)
- Area 2: Il Prodotto (5 item)
- Area 3: Il Servizio Clienti (5 item)
- Area 4: La Fattura (4 item)
- Area 5: La Soddisfazione del cliente (6 item)

Di seguito le 28 domande del questionario utilizzato:

Tab. 7.1: Questionario di Soddisfazione del cliente

Area	Domanda
Consulente	Quanto è stato cortese?
	Quanto è stato chiaro?
	Quanto è stato competente?
	Quanto è stato affidabile?
	Quanto sei stato soddisfatto dei tempi di gestione di una richiesta d'assistenza?
Prodotto	Quanto sei soddisfatto dei servizi offerti?
	Quanto sei soddisfatto del tuo canone?
	Quanto sei soddisfatto del gestore e interlocutore unico?
	Quanto sei soddisfatto del conto?
	Quanto sei soddisfatto dell'offerta ricevuta?
Servizio Clienti	Quanto è stato cortese?
	Quanto è stato chiaro?
	Quanto è stato competente?
	Quanto è stato affidabile?
	Quanto sei stato soddisfatto dei tempi di gestione di una richiesta d'assistenza?
Fattura	Quanto sei soddisfatto del rapporto qualità/prezzo?
	Quanto ritieni semplice consultare la prima pagina della fattura?
	Quanto ritieni semplice consultare la pagina del conto?
	Quanto ritieni semplice consultare i dettagli presenti in fattura?
Soddisfazione	Quanto sei soddisfatto in generale del rapporto col consulente?
	Quanto sei soddisfatto in generale del prodotto?
	Quanto sei soddisfatto in generale del Servizio Clienti?
	Quanto sei soddisfatto in generale della fattura?
	Quanto ritieni affidabile la nostra azienda?
	Quanto sei soddisfatto in generale della nostra azienda?

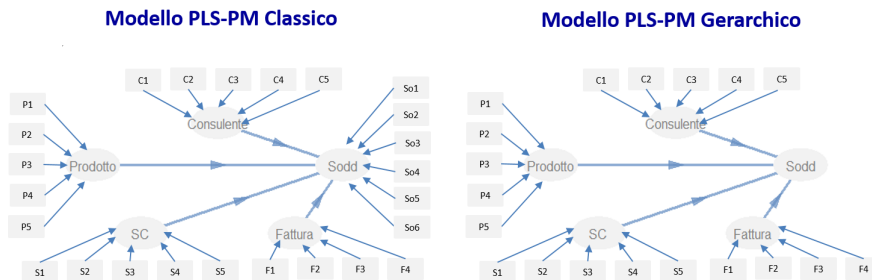
7.2.2 Il trattamento dei dati

Prima di passare a descrivere i risultati dell'analisi del modello di Customer Satisfaction è necessario soffermarsi sul trattamento dei dati. In primo luogo, dato che il pacchetto utilizzato in R⁴ per l'analisi non ammette dati mancanti, è stato necessario trattarli. Come detto in precedenza, i questionari con più del 50% dei dati mancanti sono stati eliminati, i dati mancanti relativi ai questionari rimasti sono stati imputati con il metodo del vicino più prossimo (*nearest neighbour*)⁵.

Dopo questa fase i valori di tutte le variabili sono stati riportati in una scala da 0 a 100, in primo luogo, per seguire le indicazioni fornite in letteratura [7] per il calcolo dell'ECSI utilizzando il PLS-PM, ed anche per riportare ad una stessa scala tutte le variabili prese in considerazione.

7.2.3 L'analisi del modello di customer satisfaction con PLS-PM: Validazione

Per l'analisi dei dati raccolti tramite il questionario è stato scelto l'approccio di soft-modeling PLS-PM che ci ha pertanto consentito di creare un modello adatto alle esigenze valutative e ci ha permesso di considerare assieme aspetti che si presentano in tempi diversi. In particolare i due modelli scelti per l'analisi sono:



I modelli sono stati costruiti correlando le 4 variabili latenti esogene: "Consulente", "Prodotto", "Servizio Clienti" (SC) e "Fattura" con la variabile endogena "Soddisfazione" (SODD).

⁴ <http://www.r-project.org/>.

⁵ Questo approccio è molto comune per predire i valori da imputare condizionatamente ai valori di variabili ausiliare, consiste nell'introdurre un concetto di somiglianza tra le unità, basato su un'opportuna funzione di distanza, definita sulle variabili ausiliarie. Per ciascun ricevente la scelta del donatore da cui prelevare i valori è limitata a quelle unità che minimizzano la funzione di distanza.

Ricordiamo che nella tabella 7.1 vengono riportate le variabili manifeste considerate per ogni variabile latente.

Proseguimo con l'analisi utilizzando il software R e più nello specifico il pacchetto *plspm*⁶ da cui provengono tutti gli output presenti in questo paragrafo. Tutti i comandi utilizzati in R sono riportati in allegato.

Tab. 7.2: Verifica dell'unidimensionalità dei blocchi Modello Classico.

	Modo	N.VM	C.alpha	DG.rho	1°A.v.	2°A.v.
Consulente	A	5	0,960	0,969	4,320	0,292
Prodotto	A	5	0,962	0,970	4,340	0,223
SC	A	5	0,958	0,968	4,290	0,353
Fattura	A	4	0,921	0,945	3,250	0,508
Sodd	A	6	0,927	0,943	4,410	0,561

Tab. 7.3: Verifica dell'unidimensionalità dei blocchi Modello Gerarchico.

	Modo	N.VM	C.alpha	DG.rho	1°A.v.	2°A.v.
Consulente	A	5	0,960	0,969	4,320	0,292
Prodotto	A	5	0,962	0,970	4,340	0,223
SC	A	5	0,958	0,968	4,290	0,353
Fattura	A	4	0,921	0,945	3,250	0,508
Sodd	A	5	0,946	0,964	4,240	0,748

Nelle tabelle 7.2 e 7.3 si osserva che il primo autovalore è, in tutti i casi, sufficiente a spiegare la variabilità totale del gruppo, infatti l'*Alpha di Cronbach* e il *Rho di Dillon-Goldstein* presentano valori elevati. Ciò ci indica che tutti i blocchi di variabili manifeste sono unidimensionali. Per questo tutte le variabili latenti sono state considerate come blocchi riflessivi.

Nella tabella 7.4 e nella tabella 7.5 sono indicati il peso stimato di ciascuna variabile e l'indice di comunaltà (che rappresenta la quantità di varianza che ciascuna variabile latente cattura dai suoi indicatori); esso è molto alto per quasi tutte le variabili, dunque la loro variabilità è sufficientemente spiegata dalle variabili latenti di cui sono combinazione lineare.

⁶ <http://cran.r-project.org/web/packages/plspm/index.html>.

Tab. 7.4: Tabella dei pesi e delle correlazioni per il Modello Classico

Variabili Latenti	VM	Pesi	Loading	Comunalità	Ridondanza
Consulente	Consulente1	0,192	0,902	0,814	0
	Consulente2	0,217	0,956	0,915	0
	Consulente3	0,218	0,957	0,916	0
	Consulente4	0,221	0,948	0,899	0
	Consulente5	0,229	0,881	0,775	0
Prodotto	Prodotto1	0,208	0,909	0,827	0
	Prodotto2	0,212	0,951	0,905	0
	Prodotto3	0,217	0,929	0,863	0
	Prodotto4	0,217	0,933	0,871	0
	Prodotto5	0,219	0,933	0,871	0
SC	SC1	0,217	0,849	0,721	0
	SC2	0,212	0,937	0,878	0
	SC3	0,218	0,96	0,921	0
	SC4	0,222	0,952	0,906	0
	SC5	0,211	0,929	0,863	0
Fattura	Fattura1	0,288	0,791	0,625	0
	Fattura2	0,269	0,929	0,863	0
	Fattura3	0,275	0,939	0,882	0
	Fattura4	0,281	0,938	0,879	0
Sodd	ConsulenteGen	0,193	0,831	0,690	0,642
	ProdottoGen	0,190	0,862	0,744	0,691
	SCGen	0,181	0,746	0,556	0,517
	FatturaGen	0,189	0,832	0,692	0,643
	SoddGen	0,213	0,946	0,895	0,831
	QuantoConsigli	0,199	0,915	0,837	0,778

Tab. 7.5: Tabella dei pesi e delle correlazioni per il Modello Gerarchico

Variabili Latenti	VM	Pesi	Loading	Comunalità	Ridondanza
Consulente	Consulente1	0,1983	0,903	0,816	0
	Consulente2	0,2179	0,957	0,916	0
	Consulente3	0,2198	0,958	0,917	0
	Consulente4	0,2199	0,948	0,899	0
	Consulente5	0,2203	0,877	0,769	0
Prodotto	Prodotto1	0,2073	0,909	0,827	0
	Prodotto2	0,2157	0,952	0,906	0
	Prodotto3	0,2178	0,929	0,863	0
	Prodotto4	0,2168	0,935	0,874	0
	Prodotto5	0,2152	0,935	0,874	0
SC	SC1	0,2231	0,854	0,729	0
	SC2	0,2061	0,939	0,881	0
	SC3	0,2134	0,967	0,935	0
	SC4	0,2199	0,958	0,918	0
	SC5	0,2121	0,938	0,879	0
Fattura	Fattura1	0,3046	0,806	0,650	0
	Fattura2	0,2655	0,929	0,863	0
	Fattura3	0,2706	0,938	0,880	0
	Fattura4	0,2717	0,935	0,874	0
Sodd	Y1_u1	0,2253	0,997	0,995	0,994
	Y2_u1	0,2258	0,997	0,994	0,994
	Y3_u1	0,2253	0,990	0,981	0,981
	Y3_u2	0,1068	0,514	0,264	0,264
	Y4_u1	0,2259	0,992	0,984	0,984
	Y4_u2	0,0978	0,492	0,242	0,242

Oltre a questi risultati, un altro strumento utile per controllare la validità del modello è la tabella nella quale vengono riportate le correlazioni tra le variabili manifeste e tutte le variabili latenti, per sintesi non riportiamo qui tutta la tabella (essa è riportata insieme a tutti gli output del modello). Possiamo però affermare che è confermata la validità monofattoriale in quanto tutte le variabili risultano essere correlate di più con il blocco di appartenenza che con altri blocchi. Solo la variabile *FatturaGen* del Modello Classico risulta leggermente più correlata ad una variabile latente che non è quella di appartenenza ma, siccome lo scarto è minimo, si è deciso di lasciare la variabile nel blocco *Soddisfazione* in quanto più congruente con il modello di analisi.

7.2.4 L'analisi del modello di customer satisfaction con PLS-PM: Interpretazione

In primo luogo osserviamo il diagramma del modello con i relativi impatti delle variabili latenti (in Fig. 7.4 e Fig. 7.5). Da questo è possibile capire quali sono le “dinamiche” del modello. Osserviamo come la variabile latente *Prodotto*, in entrambi gli approcci, sia quella che ha un impatto più alto sulla *Soddisfazione* (Sodd), mentre la variabile *Servizio Clienti* (SC) risulta avere un impatto abbastanza basso.

Fig. 7.4: Risultati Modello Pls Classico

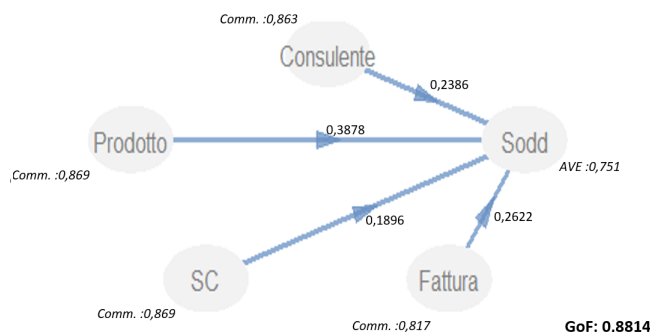
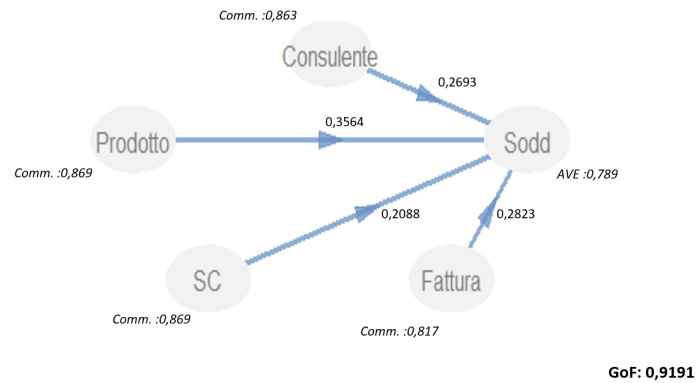


Fig. 7.5: Risultati Modello Pls Gerarchico



Risulta inoltre molto interessante notare come i due approcci diano risultati molto simili anche nelle interazioni tra le variabili latenti esogene e quella endogena.

Tab. 7.6: Punteggi medi delle variabili latenti

	Consulente	Prodotto	SC	Fattura	Soddisfazione
Media	7,166	6,806	6,743	6,650	6,587
Dev.St	2,974	2,923	2,526	2,610	2,917

Dalle tabelle 7.6 si può concludere che i clienti dell'azienda oggetto dell'analisi sono abbastanza soddisfatti, infatti i valori medi sono tutti superiori alla sufficienza. I clienti risultano particolarmente soddisfatti dell'*Consulente* mentre è proprio la *Soddisfazione* la variabile con i punteggi più bassi. È possibile inoltre andare ad analizzare la relazione che ogni singola manifesta ha sulla propria latente, in modo da capire quali sono gli aspetti sui quali si deve maggiormente puntare per migliorare il servizio.

Dalla tab. 7.4 e tab. 7.5 notiamo che quasi tutti gli impatti sono molto alti, segno che nessun aspetto è da sottovalutare, soprattutto le variabili *Competenza del Consulente* (Consulente3), *Canone* (Prodotto2), *Competenza dell'operatore* (SC4) sono aspetti su cui riporre molta attenzione, avendo impatti molto alti. Ovviamente tutti gli altri aspetti non sono comunque da sottovalutare. Di seguito (si vedano le Fig. 7.6 e Fig. 7.7) faremo considerazioni sul rapporto con le loro medie.

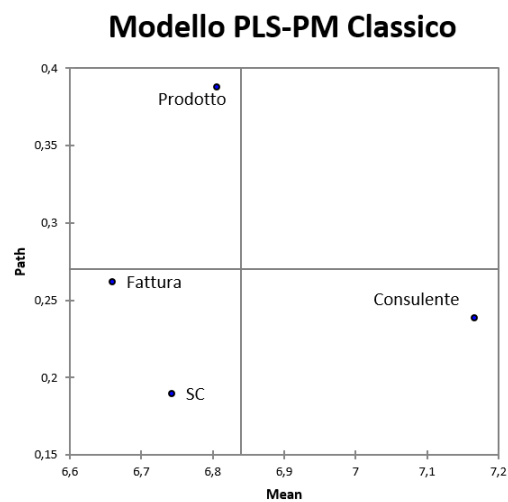


Fig. 7.6: Punteggi medi delle variabili latenti e relativi impatti sulla *Soddisfazione* con modello classico.

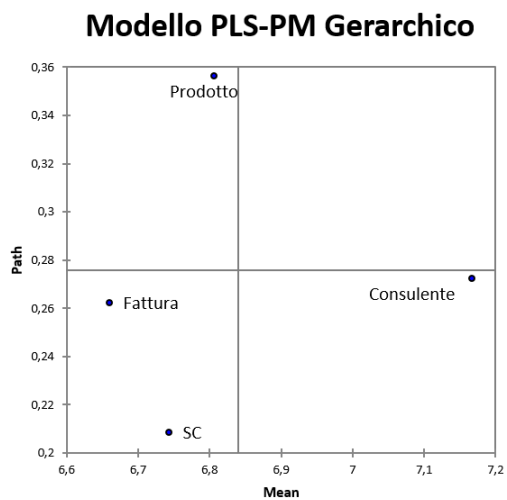


Fig. 7.7: Punteggi medi delle variabili latenti e relativi impatti sulla *Soddisfazione* con modello gerarchico.

Dato che le indagini di Customer Satisfaction sono orientate al miglioramento continuo della qualità, verificare quali aspetti debbano essere tenuti sotto controllo o eventualmente migliorati in ottica futura, risulta fondamentale. I grafici in figura 7.6 e figura 7.7 ci aiutano proprio in questo compito. Considerando contemporaneamente i punteggi medi delle variabili latenti e i relativi impatti sulla *Soddisfazione* essi ci danno l'opportunità di avere una visione di insieme dei vari aspetti. Dividendo idealmente il grafico in quadranti, potremmo definire il primo quadrante del piano cartesiano l'area di ottimo o di mantenimento, in questo quadrante si trovano quelle variabili che hanno un forte impatto sulla soddisfazione e anche una media dei punteggi alta. Nel nostro caso purtroppo nessuna variabile latente si trova in questo quadrante. Nel secondo quadrante, che potremmo definire come l'area da valutare o da valorizzare, si denota un basso impatto della variabile latente *Consulente* con una contemporanea media abbastanza alta, in questo caso la variabile latente in questione dovrebbe essere valorizzata, così da migliorarne l'impatto sulla soddisfazione. Il terzo quadrante del piano cartesiano rappresenta l'area degli aspetti da monitorare, in questo quadrante risiedono la *Fattura* e il *Servizio Clienti* (SC), trattandosi di variabili con medie dei punteggi più basse rispetto alle altre variabili, è appunto necessario monitorare e migliorare tali variabili. Infine, il quarto quadrante, anche detto area di intervento immediato, contiene la variabile *Prodotto*. In questo caso è necessario un intervento immediato in quanto la variabile ha un alto impatto sulla soddisfazione, ma una media dei punteggi bassa.

Per capire come e dove intervenire fattivamente sui vari aspetti aziendali, così da migliorare la soddisfazione, vengono utilizzati dei grafici molto simili a quelli precedenti. Incrociando le medie rilevate per ogni variabile manifesta e i relativi impatti sulle variabili latenti a cui appartengono, sarà possibile selezionare gli aspetti tangibili sui quali intervenire. Di seguito vengono visualizzati i grafici delle quattro variabili latenti considerate nel modello. I grafici sono validi per entrambi gli approcci in quanto i blocchi del primo ordine risultano essere calcolati in maniera identica.

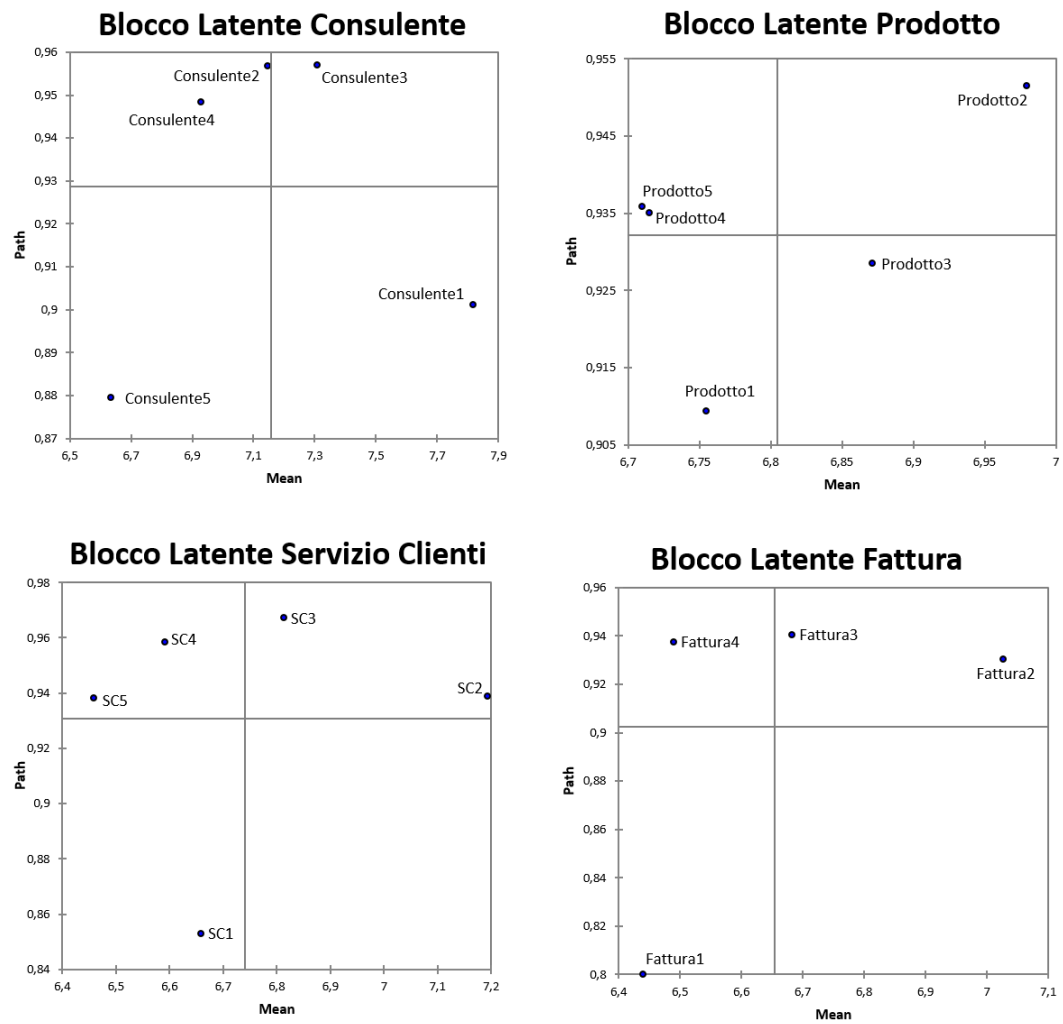


Fig. 7.8: Medie delle variabili manifeste e relativi loadings

Analizzando quindi i grafici in Fig.7.8 e focalizzandoci in primo luogo sull'area che in Fig.7.6 e Fig.7.7 era nel quadrante di intervento immediato, vediamo come per migliorare il *Prodotto*, sia necessario intervenire sul *Conto* (Prodotto4) e sull'*offerta ricevuta* (Prodotto5). Considerando ora le variabili latenti che si trovano nell'area di monitoraggio, vediamo come per migliorare la *Fattura* sia necessario in primo luogo modificare e rendere più chiari i *dettagli di consumo* (Fattura4), mentre per il *Servizio Clienti* (SC) un intervento significativo deve essere fatto verso i *tempi di gestione* (SC5) e l'*affidabilità degli operatori* (SC4). Infine se si volesse intervenire anche sull'area del *Consulente*, gli aspetti da migliorare risultano essere la *chiarezza del consulente* (Consulente2) e , anche in questo caso, la sua *affidabilità*.

7.2.5 Simulazione di applicazione

Come visto nel paragrafo precedente il confronto tra l'uso del modello gerarchico ed un modello classico ha dato risultati simili. Per capire se questo tipo di risultati può essere generalizzato si propone un primo studio simulato che, utilizzando dei dataset generati in maniera simile a come riportato nel capitolo 5⁷, calcola la percentuale di *path* ben specificati sul totale.

Come già detto, in un caso applicativo sulla customer satisfaction, le decisioni possono essere prese facendo un'analisi SWOT. L'interpretazione del grafico ottenuto è dunque il punto decisionale fondamentale. Si è deciso quindi di considerare come ben specificati i *path* che, con entrambi gli approcci, venivano riportati nello stesso quadrante del grafico, e che quindi, in termini matematici, calcolandone la differenza con la media, riportassero lo stesso segno. Le simulazioni sono state effettuate simulando un modello a 5 VL con 5 VM per blocco. Di seguito si riportano i risultati ottenuti simulando :

- 50 data set diversi con 300 unità cada uno
- 100 data set diversi con 300 unità cada uno
- 150 data set diversi con 300 unità cada uno
- 200 data set diversi con 300 unità cada uno
- 200 data set diversi con 500 unità cada uno
- 200 data set diversi con 1000 unità cada uno

⁷ l'unica differenza sta nell'aver imposto una correlazione più forte all'interno del blocco latente della variabile indipendente, così da riprodurre un blocco riflessivo

- 200 data set diversi con 1500 unità cada uno

Tab. 7.7: Percentuali di ben specificati con 5 VL e 5 VM per blocco

	N=50 R=300	N=100 R=300	N=150 R=300	N=200 R=300	N=200 R=500	N=200 R=1000	N=200 R=1500
Media	0,920	0,930	0,907	0,940	0,880	0,950	0,970
Errore St.	0,039	0,026	0,024	0,017	0,023	0,015	0,012

Come si può vedere dalla tabella 7.7, i *path* ben specificati sono quasi sempre sopra il 90%, la percentuale inoltre tende ad aumentare al crescere del numero di unità all'interno dei dataset e non sembra essere particolarmente influenzata dal numero di dataset utilizzati. Anche l'errore standard diminuisce all'aumentare della numerosità campionaria ed è sempre inferiore allo 0.05, definendo quindi una misura stabile.

7.3 Conclusioni

Partendo dagli studi sui modelli gerarchici si è proposto un ulteriore utilizzo di questi ultimi. In alcuni ambiti socio-economici, spesso, i dati sono raccolti tramite questionario. La somministrazione dei questionari risulta a volte complessa per questioni di tempo e di denaro. Diventa dunque importante ridurre la lunghezza di questo strumento di rilevazione, così da ridurre tempi, costi e tassi di abbandono. Si è presentata dunque un'applicazione del PLS-PM di alto ordine sulla Customer Satisfaction per la riduzione della lunghezza del questionario. Utilizzando un questionario classico con domande relative anche alla soddisfazione generale del cliente, si è applicato un modello PLS-PM classico contrapposto ad un modello gerarchico applicato agli stessi dati, non utilizzando le VM del blocco di alto ordine. In conclusione, osservando i risultati delle analisi PLS-PM comparate tra loro, sia da un punto di vista analitico e di validazione del modello, sia da un punto di vista interpretativo, risulta che l'utilizzo di un modello gerarchico applicato all'analisi della Customer Satisfaction porti a risultati molto simili a quelli ottenuti con il modello classico. Per convalidare questo risultato è stato anche effettuato un primo semplice studio di simulazione che ha dato risultati confortanti sulla generalizzazione di questo tipo di applicazione. Ciò ci permette dunque di utilizzare questo tipo di modello senza alterare le decisioni prese e nel contempo di ridurre la lunghezza del questionario, riducendo così tempi, costi e tasso di abbandono dell'indagine.

CONCLUSIONI

In questo lavoro di tesi è stata affrontata la questione della stima di un concetto complesso formato da aspetti diversi, ognuno dei quali rappresenta in parte il concetto preso in analisi. Numerosi fenomeni, per essere misurati, richiedono la “combinazione” di diverse dimensioni, che devono essere considerate insieme come approssimazione di quel fenomeno. Questa combinazione può essere ottenuta applicando le metodologie di creazione di indicatori compositi. La letteratura esistente offre vari approcci che permettono di ottenere un indicatore composito. In questa tesi ci siamo concentrati sulla metodologia di modellazione delle equazioni strutturali, in particolare sull’approccio modellistico a minimi quadrati parziali (PLS-PM). Tale approccio consente di stimare le relazioni causali, definite secondo un modello teorico che collega due o più concetti latenti complessi, ciascuno misurato attraverso un numero di indicatori osservabili. L’idea di base è che la complessità all’interno di un sistema può essere studiata prendendo in considerazione l’insieme delle relazioni causali tra le variabili latenti, ciascuna misurata da diverse variabili manifeste. Si è dunque discusso su alcuni miglioramenti nell’approccio alla modellizzazione PLS-PM per la stima di un sistema di indicatori compositi, soffermandoci sulla validazione di un nuovo modello con componenti gerarchiche. In un quadro di indicatori compositi, i ricercatori hanno recentemente focalizzato la loro attenzione su un aspetto particolare legato alla multidimensionalità e ad un alto livello di astrazione, ciò avviene quando un indicatore composito è molteplice e privo delle proprie variabili manifeste. In questa prospettiva si è verificato un nuovo metodo per stimare un sistema di indicatori complessi di ordine superiore, per migliorare, a livello concettuale, il significato del modello. Sono stati confrontati tre approcci: l’approccio a variabili ripetute, l’approccio misto e l’approccio con regressione PLS, investigandone il modello di misurazione e la loro predittività. Lo studio di simulazione ha preso in considerazione vari modelli semplici, composti da un numero variabile di costrutti del primo ordine che influiscono su una variabile latente di secondo ordine. Considerando la mancanza di indici globali di valutazione appropriati, si è scelto di valutare le stime con vari

indici esistenti che fossero in grado di valutare e validare varie caratteristiche del modello. Il modello PLS-R si è dimostrato dare delle buone stime in tutte le situazioni proposte. Un ulteriore aspetto affrontato in questa tesi è quello dell'applicazione dei modelli PLS-PM di alto ordine al concetto della Customer Satisfaction. Come noto, il modello teorico della Customer Satisfaction, in letteratura non viene considerato un modello gerarchico, ma la sempre più crescente necessità di snellire tale modello, ha portato all'idea di ridurre la fase di raccolta dati, sopperendo a tali informazioni proprio tramite dei modelli gerarchici. In conclusione l'applicazione dell'approccio PLS-R alla Customer Satisfaction, si è dimostrata essere funzionale, portando alla luce risultati simili a quelli ottenuti con un modello classico. Si è inoltre effettuato un primo studio simulato che sembrerebbe coadiuvare una generalizzazione di questo approccio. Facendo seguito ai risultati ottenuti, in futuro si lavorerà ad affinare l'uso di modelli gerarchici per la riduzione del questionario, esaminando più approfonditamente i risultati ottenuti, inoltre si continuerà a perfezionare lo studio dei nuovi approcci gerarchici.

APPENDICE

A. ALLEGATI

A.1 Codice R per le simulazioni

```
library(simsem)
library(plsdepot)
library(plspm)
library(boot)
library(dplyr)
library(chemometrics)
library(SimDesign)

rm(list=ls())
nblocchi=
nvar=

loading <- matrix(0, nvar, nblocchi)
rapp=nvar/nblocchi
for (i in 0:(nblocchi-1)) {
  loading[((i*rapp)+1):((i+1)*rapp)], (i+1)] <- NA
}

loading.startB <- matrix("", nvar, nblocchi)
for (i in 0:(nblocchi-1)) {
  if ((i+1)==(nblocchi)){
    loading.startB[((i*rapp)+1):((i+1)*rapp),
                    (i+1)] <- "rnorm(1,0,1)"
  } else {
    loading.startB[((i*rapp)+1):((i+1)*rapp),
                    (i+1)] <- 0.7
  }
}

loading.startA <- matrix("", nvar, nblocchi)
for (i in 0:(nblocchi-1)) {
  if ((i+1)==(1)){
    loading.startA[((i*rapp)+1):((i+1)*rapp),
                    (i+1)] <- "rnorm(1,0,1)"
  } else {
    loading.startA[((i*rapp)+1):((i+1)*rapp),
                    (i+1)] <- 0.7
  }
}
```

```

      (i+1)] <- 0.7}
    }

loading.trivial <- matrix("runif(1,-0.1,0.1)", nvar, nblocchi)
loading.trivial[is.na(loading)]<-0

LYA<- bind(loading, loading.startA,misspec=loading.trivial)
LYB<- bind(loading, loading.startB,misspec=loading.trivial)

error.cor.trivial <- matrix("rnorm(1, 0, 0.1)", nvar, nvar)
diag(error.cor.trivial) <- 1
RTE <- binds(diag(nvar), misspec=error.cor.trivial)

factor.cor <- diag(nblocchi)
for (i in 1:(nblocchi-2))
{
  for (j in (i+1):(nblocchi-1))
  {
    factor.cor[i, j] <- factor.cor[j, i] <- NA
  }
}

RPS <- binds(factor.cor, 0.01)

path <- matrix(0, nblocchi, nblocchi)
path[1:(nblocchi-1), nblocchi] <- NA
path.start <- matrix(0, nblocchi, nblocchi)

for (i in 1:(nblocchi-1))
{
  path.start[i, nblocchi] <- 0.80
}

BE <- bind(path, path.start)

pm.modelA <- model(BE=BE, LY=LYA, RPS=RPS, RTE=RTE, modelType="SEM")
pm.modelB <- model(BE=BE, LY=LYB, RPS=RPS, RTE=RTE, modelType="SEM")

dat_100A <- generate(pm.modelA, 100,drop=F)

```

```

dat_250A <- generate(pm.modelA, 250,drop=F)
dat_500A <- generate(pm.modelA, 500,drop=F)
dat_1000A <- generate(pm.modelA, 1000,drop=F)

sc = function(x){(x-min(x))/(max(x)-min(x))}
dat_100A=round(sc(dat_100)*100,digits = 0)
dat_250A=round(sc(dat_250)*100,digits = 0)
dat_500A=round(sc(dat_500)*100,digits = 0)
dat_1000A=round(sc(dat_1000)*100,digits = 0)

dat_100B <- generate(pm.modelB, 100,drop=F)
dat_250B <- generate(pm.modelB, 250,drop=F)
dat_500B <- generate(pm.modelB, 500,drop=F)
dat_1000B <- generate(pm.modelB, 1000,drop=F)

sc = function(x){(x-min(x))/(max(x)-min(x))}
dat_100B=round(sc(dat_100)*100,digits = 0)
dat_250B=round(sc(dat_250)*100,digits = 0)
dat_500B=round(sc(dat_500)*100,digits = 0)
dat_1000B=round(sc(dat_1000)*100,digits = 0)

for (i in 1:nblocchi)
{
  if (i==nblocchi)
  {
    a = paste("y",i," <-c(rep(1,(nblocchi-1)),0)",sep = "")
    eval(parse(text=a))
  }
  else
  {
    b = paste("y",i," <- c(rep(0,nblocchi))",sep="")
    eval(parse(text=b))
  }
}

tx=""
for(i in 1:nblocchi)

```

```

{
  if (i==1)
  {
    tx=paste('innerB<-rbind(y',i,sep="")
  }
  if(i != nblocchi & i != 1)
  {
    tx=paste(tx,',y',i,sep='')
  }
  if (i == nblocchi)
  {
    tx=paste(tx,',y',i,')',sep='')
  }
}
eval(parse(text=tx))

for (i in 1:nblocchi)
{
  if (i==1)
  {
    a = paste("ya",i," <- c(rep(0,nblocchi)) ",sep = "")
    eval(parse(text=a))
  }
  else
  {
    b = paste("ya",i,"<-c(1,rep(0,(nblocchi-1))) ",sep="")
    eval(parse(text=b))
  }
}

tx=""
for(i in 1:nblocchi)
{
  if (i==1)
  {
    tx=paste('innerA<-rbind(ya',i,sep="")
  }
  if(i != nblocchi & i != 1)
  {

```

```

    tx=paste(tx,',ya',i,sep='')
  }
  if (i == nblocchi)
  {
    tx=paste(tx,',ya',i,')',sep='')
  }
}
eval(parse(text=tx))

outer.t <- list()
for (i in 0:(nblocchi-1)) {
  a<-list((((i*rapp)+1):((i+1)*rapp)))
  outer.t <- append(outer.t,a)
}

SimAll<- function(inner,outer,mo, mi, data, indices,nvar,nblocchi) {
  d <- data[indices,]
  sc = function(x){(x-min(x))/(max(x)-min(x))}
  mode=rep(toupper(mo),nblocchi)
  rapp=nvar/nblocchi
  numvar=nvar-length(outer[[nblocchi]])

  if(toupper(mi)=='B'){
    outer[[nblocchi]] <- c(1:numvar)
  }
  if(toupper(mi)=='A'){
    outer[[1]] <- c((rapp+1):nvar)
  }

  res_rep<-plspm(d, inner, outer, mode
    , scheme= "centroid", scaled=T, boot.val=TRUE)

  if(toupper(mi)=='B'){
    for (i in 1:(nblocchi-1))
    {
      if (i==1)
      {
        aa = paste("d <-cbind(d,f",i

```

```

        , "=round(sc(res_rep$scores[,", i
        , "])*100,digits = 0)", sep = "")
    }
else if (i==(nblocchi-1))
{
    aa = paste(aa, ", f", i, "=round(sc(res_rep$scores[, "
        , i, "])*100,digits = 0))", sep = "")
}
else
{
    aa = paste(aa, ", f", i
        , "=round(sc(res_rep$scores[, "
        , i, "])*100,digits = 0)", sep = "")
}
}
eval(parse(text=aa))

outer[[nblocchi]] <- c((nvar+1):(nvar+(nblocchi-1)))
}

if(toupper(mi)=='A'){
  for (i in 2:(nblocchi))
  {
    if (i==2)
    {
      aa = paste("d <-cbind(d, f", i
        , "=round(sc(res_rep$scores[,", i
        , "])*100,digits = 0)", sep = "")
    }
    else if (i==(nblocchi))
    {
      aa = paste(aa, ", f", i
        , "=round(sc(res_rep$scores[,", i
        , "])*100,digits = 0))", sep = "")
    }
    else
    {
      aa = paste(aa, ", f", i
        , "=round(sc(res_rep$scores[,", i
        , "])*100,digits = 0)", sep = "")
    }
  }
}

```

```

    }
  }
  eval(parse(text=aa))
  outer[[1]] <- c((nvar+1):(nvar+(nblocchi-1)))
}

res_Mixed <- plspm(d, inner, outer, mode,
                  scheme= "centroid", scaled=T,
                  boot.val=TRUE)

if(toupper(mi)=='B'){
  for (i in 1:((dim(inner)[2])-1))
  {
    eval(parse(text=paste('T_y',i,' = plsreg2(d[,outer[[i]]],
                                          d[,c(1:numvar)], comps=NULL
                                          , crosval=TRUE)'
                          ,sep=' ')))
    eval(parse(text=paste('T_y',i,' $y.scores<-
                          as.data.frame(T_y',i
                          , '$y.scores [,which(T_y',i
                          , '$Q2[, "Q2"]> 0.09)])',sep=' ')))
    eval(parse(text=paste('if(length(T_y',i
                          , '$y.scores)==1){dimnames(T_y',i
                          , '$y.scores)[[2]]<- "u1"}',sep=' ')))
    eval(parse(text=paste('dimnames(T_y',i,
                          , '$y.scores)[[2]]<-paste("Y'
                          ,i,'_",dimnames(T_y',i
                          , '$y.scores)[[2]],sep="")',sep=' ')))

    #browser()
    if (i==1)
    {
      tx=paste('Y<-cbind(round(sc(T_y',i,sep="")
    }
    if(i != ((dim(inner)[2])-1) & i != 1)
    {
      tx=paste(tx,' $y.scores)*100,digits=0)
      , round(sc(T_y',i,sep=' '))
    }
  }
}

```



```

    if (i == ((dim(inner)[2])-1))
    {
        tx=paste(tx,'$y.scores)*100,digits=0)
                , round(sc(T_y',i
                ,'$y.scores)*100,digits=0))',sep='')
    }
}

eval(parse(text=tx))

d<-cbind(d,Y)
outer[[nblocchi]]<-c((nvar+nblocchi):
                    (nvar+nblocchi+dim(Y)[2]-1))
}

if(toupper(mi)=='A'){
  for (i in 2:((dim(inner)[2])))
  {
    eval(parse(text=paste('T_y',i,' = plsreg2(d[,outer[[i]]]
                        ,d[,c((rapp+1):nvar)], comps=NULL
                        , crosval=TRUE)',sep = '')))
    eval(parse(text=paste('T_y',i,' $y.scores<-as.data.frame(T_y',i
                        ,'$y.scores [,which(T_y',i
                        ,'$Q2[,"Q2"]> 0.09)])',sep = '')))
    eval(parse(text=paste('if(length(T_y',i
                        ,'$y.scores)==1){dimnames(T_y',i
                        ,'$y.scores)[[2]]<- "u1"}',sep = '')))
    eval(parse(text=paste('dimnames(T_y',i
                        ,'$y.scores)[[2]]<-paste("Y",i
                        ,'_",dimnames(T_y',i
                        ,'$y.scores)[[2]],sep=""',sep = '')))

    #browser()
    if (i==2)
    {
        tx=paste('Y<-cbind(round(sc(T_y',i,sep="")
    }
    if(i != ((dim(inner)[2])) & i != 2)
    {
        tx=paste(tx,'$y.scores)*100,digits=0)
                , round(sc(T_y',i,sep='')

```

```

    }
    if (i == ((dim(inner)[2])))
    {
        tx=paste(tx,'$y.scores)*100,digits=0), round(sc(T_y',i
            ,'$y.scores)*100,digits=0))',sep='')
    }
}

eval(parse(text=tx))

d<-cbind(d,Y)

outer[[1]]<-c((nvar+nblocchi):
              (nvar+nblocchi+dim(Y)[2]-1))
}

res_PLSR <- plspm(d, inner, outer, mode, scheme= "centroid"
                , scaled=T, boot.val=TRUE)

if(toupper(mi)=='A'){
  return(c(
    #repeted
    res_rep$path_coefs[-1,1]
    , res_rep$inner_summary[1,2] #r2
    , res_rep$inner_summary[1,3] #communality
    , res_rep$inner_summary[1,4] #redundancy
    , res_rep$inner_summary[1,5] #ave
    , res_rep$gof
    #Mixed
    , res_Mixed$path_coefs[-1,1]
    , res_Mixed$inner_summary[1,2] #r2
    , res_Mixed$inner_summary[1,3] #communality
    , res_Mixed$inner_summary[1,4] #redundancy
    , res_Mixed$inner_summary[1,5] #ave
    , res_Mixed$gof
    #PlsR
    , res_PLSR$path_coefs[-1,1]
    , res_PLSR$inner_summary[1,2] #r2
    , res_PLSR$inner_summary[1,3] #communality

```

```

        , res_PLSR$inner_summary[1,4] #redundancy
        , res_PLSR$inner_summary[1,5] #ave
        , res_PLSR$gof
    )))

if(toupper(mi)=='B'){
  return(c(
    #repeted
    res_rep$path_coefs[nblocchi,1:(nblocchi-1)]
    , res_rep$inner_summary[nblocchi,2] #r2
    , res_rep$inner_summary[nblocchi,3] #communality
    , res_rep$inner_summary[nblocchi,4] #redundancy
    , res_rep$inner_summary[nblocchi,5] #ave
    , res_rep$gof
    #Mixed
    , res_Mixed$path_coefs[nblocchi,1:(nblocchi-1)]
    , res_Mixed$inner_summary[nblocchi,2] #r2
    , res_Mixed$inner_summary[nblocchi,3] #communality
    , res_Mixed$inner_summary[nblocchi,4] #redundancy
    , res_Mixed$inner_summary[nblocchi,5] #ave
    , res_Mixed$gof
    #PlsR
    , res_PLSR$path_coefs[nblocchi,1:(nblocchi-1)]
    , res_PLSR$inner_summary[nblocchi,2] #r2
    , res_PLSR$inner_summary[nblocchi,3] #communality
    , res_PLSR$inner_summary[nblocchi,4] #redundancy
    , res_PLSR$inner_summary[nblocchi,5] #ave
    , res_PLSR$gof
  )))
}

results_100ModeAA <- boot(data=dat_100A,
                          inner=innerA,
                          outer =outer.t,
                          mo='a',
                          mi='a',
                          nvar=nvar,
                          nblocchi=nblocchi,

```

```

                                statistic=SimAll,
                                R=500)
RBias100AA <- bias(results_100ModeAA$t
                  , parameter = results_100ModeAA$t0
                  , type = "relative")

results_100ModeBA <- boot(data=dat_100A,
                          inner=innerA,
                          outer =outer.t,
                          mo='b',
                          mi='a',
                          nvar=nvar,
                          nblocchi=nblocchi,
                          statistic=SimAll,
                          R=500)
RBias100BA <- bias(results_100ModeBA$t
                  , parameter = results_100ModeBA$t0
                  , type = "relative")

results_100ModeBB <- boot(data=dat_100B,
                          inner=innerB,
                          outer =outer.t,
                          mo='b',
                          mi='b',
                          nvar=nvar,
                          nblocchi=nblocchi,
                          statistic=SimAll,
                          R=500)
RBias100BB <- bias(results_100ModeBB$t
                  , parameter = results_100ModeBB$t0
                  , type = "relative")

results_100ModeAB <- boot(data=dat_100B,
                          inner=innerB,
                          outer =outer.t,
                          mo='a',
                          mi='b',
                          nvar=nvar,
                          nblocchi=nblocchi,
                          statistic=SimAll,

```

```

R=500)
RBias100AB <- bias(results_100ModeAB$t
, parameter = results_100ModeAB$t0
, type = "relative")

results_250ModeAA <- boot(data=dat_250A,
inner=innerA,
outer =outer.t,
mo='a',
mi='a',
nvar=nvar,
nblocchi=nblocchi,
statistic=SimAll,
R=500)
RBias250AA <- bias(results_250ModeAA$t
, parameter = results_250ModeAA$t0
, type = "relative")

results_250ModeBA <- boot(data=dat_250A,
inner=innerA,
outer =outer.t,
mo='b',
mi='a',
nvar=nvar,
nblocchi=nblocchi,
statistic=SimAll,
R=500)
RBias250BA <- bias(results_250ModeBA$t
, parameter = results_250ModeBA$t0
, type = "relative")

results_250ModeBB <- boot(data=dat_250B,
inner=innerB,
outer =outer.t,
mo='b',
mi='b',
nvar=nvar,
nblocchi=nblocchi,
statistic=SimAll,
R=500)

```

```

RBias250BB <- bias(results_250ModeBB$t
                  , parameter = results_250ModeBB$t0
                  , type = "relative")

results_250ModeAB <- boot(data=dat_250B,
                          inner=innerB,
                          outer =outer.t,
                          mo='a',
                          mi='b',
                          nvar=nvar,
                          nblocchi=nblocchi,
                          statistic=SimAll,
                          R=500)
RBias250AB <- bias(results_250ModeAB$t
                  , parameter = results_250ModeAB$t0
                  , type = "relative")

results_500ModeAA <- boot(data=dat_500A,
                          inner=innerA,
                          outer =outer.t,
                          mo='a',
                          mi='a',
                          nvar=nvar,
                          nblocchi=nblocchi,
                          statistic=SimAll,
                          R=500)
RBias500AA <- bias(results_500ModeAA$t
                  , parameter = results_500ModeAA$t0
                  , type = "relative")

results_500ModeBA <- boot(data=dat_500A,
                          inner=innerA,
                          outer =outer.t,
                          mo='b',
                          mi='a',
                          nvar=nvar,
                          nblocchi=nblocchi,
                          statistic=SimAll,
                          R=500)
RBias500BA <- bias(results_500ModeBA$t

```

```

, parameter = results_500ModeBA$t0
, type = "relative")

results_500ModeBB <- boot(data=dat_500B,
                          inner=innerB,
                          outer =outer.t,
                          mo='b',
                          mi='b',
                          nvar=nvar,
                          nblocchi=nblocchi,
                          statistic=SimAll,
                          R=500)
RBias500BB <- bias(results_500ModeBB$t
, parameter = results_500ModeBB$t0
, type = "relative")

results_500ModeAB <- boot(data=dat_500B,
                          inner=innerB,
                          outer =outer.t,
                          mo='a',
                          mi='b',
                          nvar=nvar,
                          nblocchi=nblocchi,
                          statistic=SimAll,
                          R=500)
RBias500AB <- bias(results_500ModeAB$t
, parameter = results_500ModeAB$t0
, type = "relative")

results_1000ModeAA <- boot(data=dat_1000A,
                          inner=innerA,
                          outer =outer.t,
                          mo='a',
                          mi='a',
                          nvar=nvar,
                          nblocchi=nblocchi,
                          statistic=SimAll,
                          R=500)
RBias1000AA <- bias(results_1000ModeAA$t
, parameter = results_1000ModeAA$t0

```

```

, type = "relative")

results_1000ModeBA <- boot(data=dat_1000A,
                           inner=innerA,
                           outer =outer.t,
                           mo='b',
                           mi='a',
                           nvar=nvar,
                           nblocchi=nblocchi,
                           statistic=SimAll,
                           R=500)
RBias1000BA <- bias(results_1000ModeBA$t
, parameter = results_1000ModeBA$t0
, type = "relative")

results_1000ModeBB <- boot(data=dat_1000A,
                           inner=innerB,
                           outer =outer.t,
                           mo='b',
                           mi='b',
                           nvar=nvar,
                           nblocchi=nblocchi,
                           statistic=SimAll,
                           R=500)
RBias1000BB<- bias(results_1000ModeBB$t
, parameter = results_1000ModeBB$t0
, type = "relative")

results_1000ModeAB <- boot(data=dat_1000B,
                           inner=innerB,
                           outer =outer.t,
                           mo='a',
                           mi='b',
                           nvar=nvar,
                           nblocchi=nblocchi,
                           statistic=SimAll,
                           R=500)
RBias1000AB <- bias(results_1000ModeAB$t
, parameter = results_1000ModeAB$t0
, type = "relative")

```


A.2 Codice R per le simulazioni sull'Applicazione

```
library(simsem)
library(plsdepot)
library(plspm)
library(boot)
library(dplyr)
library(boot)
library(chemometrics)
library(SimDesign)

rm(list=ls())
nblocchi=5
nvar=25

loading <- matrix(0, nvar, nblocchi)
rapp=nvar/nblocchi
for (i in 0:(nblocchi-1)) {
  loading[((i*rapp)+1):((i+1)*rapp)], (i+1)] <- NA
}
loading.start <- matrix("", nvar, nblocchi)
for (i in 0:(nblocchi-1)) {
  if ((i+1)==(nblocchi)){
    loading.start[((i*rapp)+1):((i+1)*rapp),
                  (i+1)] <- "rnorm(1,0.5,0.5)"
  } else {
    loading.start[((i*rapp)+1):((i+1)*rapp),
                  (i+1)] <- 0.7
  }
}

loading.trivial <- matrix("runif(1,-0.1,0.1)",
                          nvar, nblocchi)
loading.trivial[is.na(loading)]<-0

LY<- bind(loading, loading.start,misspec=loading.trivial)

error.cor.trivial <- matrix("rnorm(1, 0, 0.1)",
                             nvar, nvar)
```

```

diag(error.cor.trivial) <- 1
RTE <- binds(diag(nvar), misspec=error.cor.trivial)

factor.cor <- diag(nblocchi)
for (i in 1:(nblocchi-2))
{
  for (j in (i+1):(nblocchi-1))
  {
    factor.cor[i, j] <- factor.cor[j, i] <- NA
  }
}

RPS <- binds(factor.cor, 0.01)

path <- matrix(0, nblocchi, nblocchi)
path[1:(nblocchi-1), nblocchi] <- NA
path.start <- matrix(0, nblocchi, nblocchi)

for (i in 1:(nblocchi-1))
{
  path.start[i, nblocchi] <- 0.80
}

BE <- bind(path, path.start)

pm.model <- model(BE=BE, LY=LY, RPS=RPS, RTE=RTE, modelType="SEM")

for (i in 1:nblocchi)
{
  if (i==nblocchi)
  {
    a = paste("y",i," <-c(rep(1,(nblocchi-1)),0)",sep = "")
    eval(parse(text=a))
  }
  else
  {
    b = paste("y",i," <- c(rep(0,nblocchi))",sep="")
    eval(parse(text=b))
  }
}

```

```

tx=""
for(i in 1:nblocchi)
{
  if (i==1)
  {
    tx=paste('innerB<-rbind(y',i,sep="")
  }
  if(i != nblocchi & i != 1)
  {
    tx=paste(tx,',y',i,sep='')
  }
  if (i == nblocchi)
  {
    tx=paste(tx,',y',i,')',sep='')
  }
}
eval(parse(text=tx))

for (i in 1:nblocchi)
{
  if (i==1)
  {
    a = paste("ya",i,"<- c(rep(0,nblocchi)) ",sep = "")
    eval(parse(text=a))
  }
  else
  {
    b = paste("ya",i,"<-c(1,rep(0,(nblocchi-1))) ",sep="")
    eval(parse(text=b))
  }
}

outer.t <- list()
for (i in 0:(nblocchi-1)) {
  a<-list((((i*rapp)+1):((i+1)*rapp)))
  outer.t <- append(outer.t,a)
}

```

```

SimRVsClass<- function(inner,outer,mo,data, ind,nvar,nblocchi,pmm) {
  ResTot = list()
  for(un in c(1:ind)){
    outeruse <- outer
    pm.model <- pmm
    dat_d <- generate(pm.model, floor(runif(1,(data-50),(data+50))),drop=F)
    sc = function(x){(x-min(x))/(max(x)-min(x))}
    dataD=round(sc(dat_d)*100,digits = 0)
    mode=rep(toupper(mo),nblocchi)
    rapp=nvar/nblocchi
    numvar=nvar-length(outeruse[[nblocchi]])

    res_Classic<-plspm(dataD, inner, outeruse, mode
                        , scheme= "centroid", scaled=T, boot.val=TRUE)

    outeruse[[nblocchi]] <- c(1:numvar)

    for (i in 1:((dim(inner)[2])-1))
    {
      eval(parse(text=paste('T_y',i,' = plsreg2(dataD[,outeruse[[i]]],
                                dataD[,c(1:numvar)], comps=NULL, crosval=TRUE)'
                                , sep ='')))
      eval(parse(text=paste('T_y',i,' $y.scores<-as.data.frame(T_y',i
                                , '$y.scores [,which(T_y',i
                                , '$Q2[, "Q2"> 0.09)])', sep='')))
      eval(parse(text=paste('if(length(T_y',i
                                , '$y.scores)==1){dimnames(T_y',i
                                , '$y.scores)[[2]]<- "u1"}', sep='')))
      eval(parse(text=paste('dimnames(T_y',i,' $y.scores) [[2]]<-paste("Y'
                                , i, '_"',dimnames(T_y',i
                                , '$y.scores) [[2]], sep="")', sep='')))

      if (i==1)
      {
        tx=paste('Y<-cbind(round(sc(T_y',i, sep="")
      }
      if(i != ((dim(inner)[2])-1) & i != 1)
      {
        tx=paste(tx,' $y.scores)*100,digits=0), round(sc(T_y',i, sep='')
      }
    }
  }
}

```

```

    if (i == ((dim(inner)[2])-1))
    {
        tx=paste(tx,'$y.scores)*100,digits=0), round(sc(T_y',i
            ,'$y.scores)*100,digits=0))',sep='')
    }
}

eval(parse(text=tx))

dataD<-cbind(dataD,Y)
outeruse[[nblocchi]]<-c((nvar+1):(nvar+1+dim(Y)[2]-1))
res_PLSR <- plspm(dataD, inner, outeruse, mode, scheme= "centroid"
    , scaled=T, boot.val=TRUE)

res_Classic$path_coefs[-1,1]-(mean(res_Classic$path_coefs[-1,1]))
res_PLSR$path_coefs[-1,1]-(mean(res_PLSR$path_coefs[-1,1]))

PercEntage=sum((sign(res_Classic$path_coefs[-1,1]-
    (mean(res_Classic$path_coefs[-1,1])))==
    sign(res_PLSR$path_coefs[-1,1]-
    (mean(res_PLSR$path_coefs[-1,1]))))*1)/
    (nblocchi-1)
ResTot[[un]]<-PercEntage
}

return(unlist(ResTot))

}

results_50_300 <- SimRVsClass(data=300,
    inner=innerB,
    outer =outer.t,
    mo='a',
    nvar=nvar,
    nblocchi=nblocchi,
    pmm=pm.model,
    ind = 50)

results_100_300 <- SimRVsClass(data=300,
    inner=innerB,

```

```

        outer =outer.t,
        mo='a',
        nvar=nvar,
        nblocchi=nblocchi,
        pmm=pm.model,
        ind = 100)

results_150_300 <- SimRVsClass(data=300,
        inner=innerB,
        outer =outer.t,
        mo='a',
        nvar=nvar,
        nblocchi=nblocchi,
        pmm=pm.model,
        ind = 150)

results_200_300 <- SimRVsClass(data=300,
        inner=innerB,
        outer =outer.t,
        mo='a',
        nvar=nvar,
        nblocchi=nblocchi,
        pmm=pm.model,
        ind = 200)

results_200_500 <- SimRVsClass(data=500,
        inner=innerB,
        outer =outer.t,
        mo='a',
        nvar=nvar,
        nblocchi=nblocchi,
        pmm=pm.model,
        ind = 200)

results_200_1000 <- SimRVsClass(data=1000,
        inner=innerB,
        outer =outer.t,
        mo='a',
        nvar=nvar,
        nblocchi=nblocchi,

```

```

                                pmm=pm.model,
                                ind = 200)

results_200_1500 <- SimRVsClass(data=1500,
                                inner=innerB,
                                outer =outer.t,
                                mo='a',
                                nvar=nvar,
                                nblocchi=nblocchi,
                                pmm=pm.model,
                                ind = 200)

MedieRes<-c(mean(results_50_300)
,mean(results_100_300)
,mean(results_150_300)
,mean(results_200_300)
,mean(results_200_500)
,mean(results_200_1000)
,mean(results_200_1500))

std<-function(x){sd(x)/sqrt(length(x))}

ErrorRes<-c(
  std(results_50_300)
, std(results_100_300)
, std(results_150_300)
, std(results_200_300)
, std(results_200_500)
, std(results_200_1000)
, std(results_200_1500))

```

BIBLIOGRAFIA

- [1] Addinsoft. «**XLSTAT** - Statistics Package for **Excel**». In: URL <http://www.xlstat.com/> (2011).
- [2] R. Agarwal e E. Karahanna. «Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage». In: *MIS Quarterly*, 24 (4), 665 - 694 (2000).
- [3] E. Amaturio. *Metodologia della ricerca sociale*. UTET università, 2012.
- [4] J.C. Anderson e D.W. Gerbing. «Structural equation modeling in practice: A review and recommended two-step approach». In: *Psychological Bulletin*, 103, 411 - 423 (1988).
- [5] R. J. Ball. «The significance of simultaneous methods of parameter estimation in econometric models». In: *Applied Statistics*, 12, 14 - 25 (1963).
- [6] R.M. Baron e D.A. Kenny. «The Moderator Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations». In: *Journal of Personality and Social Psychology*, 51, (6) 1173 - 1182 (1986).
- [7] M.P. Bayol et al. «Use of PLS Path Modeling to Estimate the European Consumer Satisfaction Index (ECSI) Model». In: *Statistica Applicata - Italian Journal of Applied Statistics*, 12 (3), 361 - 375 (2000).
- [8] J.M. Becker, K. Klein e M. Wetzels. «Hierarchical Latent Variable Models in PLS-SEM: Guidelines for Using Reflective-Formative Type Models». In: *Long Range Planning*, 45(5-6), 363 - 365 (2012).
- [9] J. P. Benzécri. «Pratique de L'Analyse des Données, Analyse Des Correspondances». In: *Exposé Élémentaire*, Dunod, Bordas, Paris (1980).
- [10] B. Berelson e P. F. Lazarsfeld. *The Analysis of Communication Content*. University of Chicago e Columbia University, 1952.
- [11] L. von Bertalanffy. *General System Theory: Foundations, Development, Applications*. Braziller, G. (Ed), 1968.

- [12] R. Boudon e P. F. Lazarsfeld. *L'analisi empirica nelle scienze sociali*. Il Mulino, 1969.
- [13] N.M Bradburn, S. Sudman e B. Wansink. *Asking Questions: The Definitive Guide to Questionnaire Design - For Market Research, Political Polls, and Social and Health Questionnaires*. John Wiley e Sons, 2004.
- [14] E. Bruti Liberati e M. Fortis. *Le imprese multiutility. Aspetti generali e prospettive dei settori a rete*. Il Mulino, 2001.
- [15] A. Bryman e B. Burgess. *Analyzing Qualitative Data*. Taylor & Francis, 2002.
- [16] R. S. Burt. «Confirmatory Factor-Analytic Structures and the Theory Construction Process». In: *Sociological Methods and Research*, 45, 131 - 190 (1973).
- [17] E. Caldani. *Marketing strategico*. Etas Libri, 1995.
- [18] D.T. Campbell. «Qualitative knowing in action research». In: *The social context of methods*, 184 - 209 (1978).
- [19] T.A. Carte e C. J. Russell. «In pursuit of moderation: Nine common errors and their solution». In: *MIS Quarterly*, 27, 479 - 501 (2003).
- [20] R. Cataldo et al. «Developments in Higher-Order PLS-PM for the building of a system of Composite Indicators». In: *Quality & Quantity*, 51 (2), 657 - 674 (2017).
- [21] L. Cherchye et al. «One market, one number? A composite indicator assessment of EU internal market dynamics». In: *Internal Paper of the Knowledge Economy Indicators Project*. paper. University of Leuven, 2005.
- [22] L. Cherchye et al. «One Market, One Number? A Composite Indicator Assessment of EU Internal Market Dynamics». In: *European Economic Review* (2006).
- [23] W.W. Chin. *Bootstrap cross-validation indices for pls path model assessment*. Esposito Vinzi, V. et al., 2010.
- [24] W.W. Chin. *Frequently Asked Questions-Partial Least Squares and PLS-Graph*. <http://discnt.cba.uh.edu/chin/plsfac.htm>, 2000.
- [25] W.W. Chin. *How to write up and report PLS analyses*. Esposito Vinzi, V. et al., 2010.
- [26] W.W. Chin. «Issues and opinion on structural equation modeling». In: *MIS Quarterly*, 22 (1), vii - xvi (1998).

- [27] W.W. Chin. *The Partial Least Squares Approach to Structural Equation Modeling*. Marcoulides G. A. Editor, *Modern Business Research Methods*, Mahwah, NJ: Lawrence Erlbaum Associates, 295 - 336, 1998.
- [28] W.W. Chin e J. Dibbern. *An Introduction to a Permutation Based Procedure for Multi-Group PLS Analysis: Results of Tests of Differences on Simulated Data and a Cross Cultural Analysis of the Sourcing of Information System Services between Germany and the USA*. Esposito Vinzi, V. et al., 2010.
- [29] W.W. Chin e J. Dibbern. *The partial least squares approach for structural equation modeling*. Marcoulides, G.A. (Ed), *Modern methods for business research*, London: Lawrence Erlbaum Associates ,295 - 336, 2007.
- [30] W.W. Chin e A. Gopal. «Adoption intention in GSS: Relative importance of beliefs». In: *The Data Base for Advances in Information Systems*, 26 (2&3), 42 - 64 (1995).
- [31] W.W. Chin, B.L. Marcolin e P.N. Newsted. «A partial least squares approach for measuring interaction effects: results from a Monte Carlo simulation study and an electronic mail emotion/adoption study». In: *Information Systems Research*, 14 (2), 189 - 217 (2003).
- [32] G.C. Chow. «Tests of Equality between Sets of Coefficients in Two Linear Regressions». In: *Econometrica*, 28 (3), 591 - 605 (1960).
- [33] E. Ciavolino. «General distress as second order latent variable estimated through PLS-PM approach». In: *Electronic Journal of Applied Statistical Analysis*, 5 (3) 458- 464 (2013).
- [34] E. Ciavolino, M Carpita e M. Nitti. «High-order pls path model with qualitative external information». In: *Quality & Quantity*, 49(4) 1609-1620 (2015).
- [35] E. Ciavolino e M. Nitti. «A deflated indicators approach for estimating second-order reflective models through PLS-PM: an empirical illustration». In: *Journal of Applied Statistics*, 41 (10) 2222- 2239 (2014).
- [36] E. Ciavolino e M. Nitti. «Simulation study for PLS path modelling with high-order construct: A job satisfaction model evidence». In: *Advanced Dynamic Modeling of Economic and Social Systems*, 185- 207 (2013).
- [37] E. Ciavolino e M. Nitti. «Using the hybrid two-step estimation approach for the identification of second-order latent variable models». In: *Journal of Applied Statistics*, 40 (3) 508- 526 (2013).

- [38] E. Ciavolino et al. «High-order PLS path model for multi-group analysis: the prosumership service quality model». In: *Quality & Quantity*, 1- 14 (2018).
- [39] J. Cohen. *Statistical power analysis for the behavioral sciences*. Mahwah, NJ: Lawrence Erlbaum, 1988.
- [40] S. De Jong. «PLS shrinks». In: *Journal of Chemometrics*, 9 (4), 323 - 326 (1995).
- [41] P. De Muro, M. Mazziotta e A. Pareto. «Composite indices of development and poverty: An application to MDG Indicators». In: *Social indicators research*, 104 (1), 1 - 18 (2011).
- [42] F. Del Vecchio. *Scale di misura e indicatori sociali*. Cacucci, 1995.
- [43] A. Diamantopoulos e H.M. Winklhofer. «Index construction with formative indicators: An alternative to scale development». In: *Journal of Marketing Research*, 38 (2), 269 - 277 (2001).
- [44] P. Dickens, M. Valentova e M. Borsenberger. «A multidimensional assessment of social cohesion in 47 European countries». In: *CEP/INSTEAD, Luxembourg* (2011).
- [45] U. Ebert e H. Welsch. «Meaningful environmental indices: a social choice approach». In: *Journal of Environmental Economics and Management*, 47, 270 - 283 (2004).
- [46] J.R. Edwards. «Multidimensional constructs in organizational behavior research: an integrative analytical framework». In: *Organizational Research Methods*, 4 (2), 144 - 192 (2001).
- [47] J.R. Edwards e R. P. Bagozzi. «On the Nature and Direction of the Relationship between Constructs and Measures». In: *Psychological Methods*, 5 (2), 155 - 174 (2000).
- [48] V. Esposito Vinzi e C.N. Lauro. «PLS regression and classification». In: *Proceedings of the PLS'03 International Symposium*, 45 - 56. paper. France, 2003.
- [49] V. Esposito Vinzi e G. Russolillo. *Partial least squares path modeling and regression*. Wegman, Y., Scott, D. (Eds), *Wiley Interdisciplinary Reviews: Computational Statistics*, John Wiley and Sons, 2010.
- [50] V. Esposito Vinzi, L. Trinchera e S. Amato. *PLS path modeling: From Foundations to Recent developments and open issues for model assessment and improvement*. Esposito Vinzi, V. et al., 2010.

- [51] V. Esposito Vinzi et al. «PLS Path Modeling: Some Recent Methodological Developments, a Software Integrated in XLSTAT and Its Application to Customer Satisfaction Studies». In: *Proceedings of the Academy of Marketing Science Conference Marketing Theory and Practice in an Inter-Functional World*. paper. Verona, Italy, 2007.
- [52] V. Esposito Vinzi et al. «REBUS-PLS: A response-based procedure for detecting unit segments in PLS path modeling». In: *Applied Stochastic Models in Business and Industry*, 24 (5), 439 - 458 (2008).
- [53] C. Fornell e F. L. Bookstein. «Two structural equation models: LISREL and PLS applied to consumer exit-voice theory». In: *Journal of Marketing Research*, 19 (4), 440 - 452 (1982).
- [54] C. Fornell e Larcker. «Evaluating structural equation models with unobservable variables and measurement error». In: *Journal of Marketing Research*, 18, 39 - 50 (1981).
- [55] S.O. Funtowicz e J.R. Ravetz. «Uncertainty and Quality in Science for Policy». In: Kluwer Academic Publishers, *Dordrecht, Nederland* (1990).
- [56] S.O. Funtowicz e J.R. Ravetz. *Uncertainty and Quality in Science for Policy*. Kluwer Academic Publishers, Dordrecht, NL, 1990.
- [57] S. Geisser. «The predictive sample reuse method with applications». In: *Journal of the American Statistical Association*, 70, 320 - 328 (1975).
- [58] A. Gifi. *Nonlinear Multivariate Analysis*. Wiley, Chichester, UK, 1990.
- [59] I. Gonzalez Ortega, E. Echeburua e P. De Corral. «Variables significativas en las relaciones violentas en parejas jóvenes: una revisión». In: *Behavioral Psychology*, 16 (2), 207 - 225 (2008).
- [60] R. L. Gorsuch. *Factor Analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1983.
- [61] G. Guido, F. Bassi e A. Peluso. *La soddisfazione del consumatore: la misura della Customer Satisfaction nelle esperienze di consumo*. FrancoAngeli, 2010.
- [62] C. Guinot, J. Latreille e M. Tenenhaus. «PLS path modelling and multiple table analysis. Application to the cosmetic habits of women in Ile-de-France». In: *Chemometr. Intell. Lab. Syst.*, 58 (2), 247 - 259 (2001).

- [63] C. Hahn et al. «Capturing customer heterogeneity using a finite mixture PLS approach». In: *Schmalenbach Business Review*, 54, 243 – 269 (2002).
- [64] J.F. Hair, C.M. Ringle e M. Sarstedt. «PLS-SEM: Indeed a silver bullet». In: *Journal of Marketing Theory and Practice*, 19, 139 – 151 (2011).
- [65] J.F. Hair et al. *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. Thousand Oaks, CA: Sage, 2014.
- [66] J.F. Hair et al. «An assessment of the use of partial least squares structural equation modeling in marketing research». In: *Journal of the Academy of Marketing Science*, 40, 414 – 433 (2011).
- [67] J. Hall. «Measuring Progress – An Australian Travelogue». In: *Journal of Official Statistics*, 21 (4), 727 - 746 (2005).
- [68] M. Hanafi. «PLS path modelling: computation of latent variables with the estimation mode B». In: *Comput Stat*, 22 (2), 275 - 292 (2007).
- [69] S. Helm, A. Eggert e I. Garnefeld. *Modelling the impact of corporate reputation on customer satisfaction and loyalty using PLS*. Esposito Vinzi, V. et al., 2010.
- [70] J. Henseler. «On the convergence of the partial least squares path modeling algorithm». In: *Computational Statistics*, 25 (1), 107 - 120 (2010).
- [71] J. Henseler e W.W. Chin. «A comparison of approaches for the analysis of interaction effects between latent variables using partial least squares path modeling». In: *Structural Equation Modeling*, (17), 82 - 109 (2010).
- [72] J. Henseler e G. Fassott. *Testing moderating effects in PLS path models: An illustration of available procedures*. Esposito Vinzi, V. et al., 2010.
- [73] J. Henseler, C.M. Ringle e R.R. Sinkovics. *The use of partial least squares path modeling in international marketing*. Sinkovics, R.R. e Ghauri, P.N. (Eds), *Advances in International Marketing*, Emerald, Bingley, 277 - 319, 2009.
- [74] J. Henseler e M. Sarstedt. «Goodness-of-fit indices for partial least squares path modeling». In: *Computational Statistics*, 28 (2) (2013).

- [75] C. Hock, C.M. Ringle e M. Sarstedt. «Management of multipurpose stadiums: Importance and performance measurement of service interfaces». In: *International Journal of Services Technology and Management*, 14, 188 - 207 (2010).
- [76] M. Hoerger. «Participant Dropout as a Function of Survey Length in Internet-Mediated University Studies: Implications for Study Design and Voluntary Participation in Psychological Research». In: *Cyberpsychology, Behavior and Social Networking*, 13 (6): 697 - 700 (2010).
- [77] K. Holzinger e F. Swineford. «The bi-factor method». In: *Psychometrika*, 2 (1), 41 - 54 (1937).
- [78] J. Hulland, M.J. Ryan e R.K. Rayner. *Modeling customer satisfaction: a comparative performance evaluation of covariance structure analysis versus partial least squares*. Esposito Vinzi, V. et al., 2010.
- [79] R. Jacobs, P. Smith e M. Goddard. «Measuring performance: an examination of composite performance indicators». In: *Centre for Health Economics*, 29 (2004).
- [80] W. James. *The Meaning of Truth*. 1909.
- [81] D. Jarvis, S. MacKenzie e P. Podsakoff. «A critical review of construct indicators and measurement model misspecification in marketing and consumer research». In: *Journal of Consumer Research*, 30 (3), 199 - 218 (2003).
- [82] K. Jedidi, H.S. Jagpal e W.S. DeSarbo. «STEMM: A general finite mixture structural equation model». In: *Journal of Classification*, 14 (1), 23 - 50 (1997).
- [83] R.E. Johnson et al. «Recommendations for improving the construct clarity of higher-order multidimensional constructs». In: *Human Resource Management Review*, 22 (2), 67 - 72 (2012).
- [84] K.G. Jöreskog. «A general method for analysis of covariance structure». In: *Biometrika*, 57, 239 - 251 (1970).
- [85] K.G. Jöreskog. «Simultaneous factor analysis in several populations». In: *Psychometrika*, 57, 409 - 426 (1971).
- [86] K.G. Jöreskog. «Structural analysis of covariance and correlation matrices». In: *Psychometrika*, 43, 443 - 477 (1978).

- [87] K.G. Jöreskog. *Testing structural equation models*. Bolle, K.A. e Long, J.S. (Eds) *Testing Structural Equation Models*, Sage Publication, Newbury Park, 1993.
- [88] K.G. Jöreskog e M. van Thillo. «LISREL: A general computer program for estimating a linear structural equation system involving multiple indicators of unmeasured variables». In: ETS Research Bulletin Series, 2, 1 - 71 (1972).
- [89] K.G. Jöreskog e H. Wold. *The ML and PLS techniques for modeling with latent variables: historical and comparative aspects*. Joreskog, K.G. e Wold, H. (Eds) *Systems Under Indirect Observation*, Part 1, North-Holland, Amsterdam, 263 – 270, 1982.
- [90] F. Joseph et al. *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. SAGE Publications, Inc., 2014.
- [91] H.F. Kaiser. «The varimax criterion for analytic rotation in factor analysis». In: Psychometrika, 23 (3), 187 - 200 (1958).
- [92] M. Keil et al. «A cross-cultural study on escalation of commitment behavior in software projects». In: MIS Quarterly, 24, 299 - 325 (2000).
- [93] G. Kim, B. Shin e V. Grover. «Investigating two contradictory views of formative measurement in information systems research». In: MIS Quarterly, 34, 345 – 365 (2010).
- [94] R. Kline. *Principles and practice of structural equation modeling*. Guilford Press, New York, 1998.
- [95] B.R. Koka e J.E. Prescott. «Strategic alliances as social capital: a multidimensional view». In: Strategic Management Journal, 23 (9), 795 - 816 (2002).
- [96] N. Krámer. «Analysis of high-dimensional data with partial least squares and boosting». Tesi di dott. Technischen Universität Berlin, Berlin, Germany, 2007.
- [97] J. Kruskal. «Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis». In: Psychometrika, 29 (2), 115 - 129 (1964).
- [98] K.S. Law e C. Wong. «Multidimensional Constructs In Structural Equation Analysis: An Illustration Using the Job Perception and Job Satisfaction Constructs». In: Journal of Management, 25 (2), 143 - 160 (1999).

- [99] K.S. Law, C. Wong e W.H. Mobley. «Toward a Taxonomy of Multi-dimensional Constructs». In: *Academy of Management Review*, 23 (4), 741 - 755 (1998).
- [100] P. F. Lazarsfeld. *Problems in Methodology*, in: *Sociology Today: Problems and Prospects*. Basic Books, New York, 1958.
- [101] P. F. Lazarsfeld e R. Henry. *Latent Structure Analysis*. Houghton Mifflin, Boston, 1968.
- [102] L. Lebart, A. Morineau e Fénelon J.P. *Traitement des données statistiques*. 1979.
- [103] L. Lee et al. «On the use of partial least squares path modeling in accounting research». In: *International of Accounting Information Systems* (2011).
- [104] J. de Leeuw e J. Van Rijckevorsel. *HOMALS and PRINCALS. Some generalizations of principal components analysis*. Diday, E. e et al. (Eds), *Data Analysis and Informatics*, Amsterdam: North-Holland, 1980.
- [105] J. de Leeuw, F.W. Young e Y. Takane. «Additive structure in qualitative data: an alternating least squares method with optimal scaling features». In: *Psychometrika*, 41, 471 - 503 (1976).
- [106] T.D. Little, J.A. Bovaird e K.F. Widaman. «On the merits of orthogonalizing powered and product terms: Implications for modeling latent variable interactions». In: *Structural Equation Modeling* (2006).
- [107] J.B. Lohmöller. *Latent Variable Path Modeling with Partial Least Squares*. Physica, Verlag, Heidelberg, Germany, 1989.
- [108] J.B. Lohmöller. *PLS-PC: Latent Variables Path Analysis with Partial Least Squares - Version 1.8 for PCs under MS-Dos*. 1987.
- [109] Freudenberg M. «Composite indicators of country performance: a critical assessment». In: *OECD, Paris* (2003).
- [110] B. MacElroy. «Variables influencing dropout rates in Web-based surveys». In: *Quirk's Marketing Research*, 0605 (2000).
- [111] S.B. MacKenzie, P.M. Podsakoff e C.B. Jarvis. «The Problem of Measurement Model Misspecification in Behavioral and Organizational Research and Some Recommended Solutions». In: *Journal of Applied Psychology*, 90 (4), 710 - 730 (2005).
- [112] T. Mannarini et al. «The role of affects in culture-based interventions: Implications for practice». In: *Psychology*, 3 (8), 569 - 577 (2012).

- [113] A. Marradi. «Classificazioni, Tipologie, Tassonomie». In: *Enciclopedia delle Scienze Sociali*, vol. II, 22 - 30 (1993).
- [114] A. Marradi. «Fedeltà di un dato, affidabilità di una definizione operativa». In: *Rassegna Italiana di Sociologia*, XXXI (1), 55 - 96 (1990).
- [115] A. Marradi. *Metodologia delle scienze sociali*. Il Mulino, Bologna, 2007.
- [116] A. Marradi. «Misurazione e scale: qualche riflessione e una proposta». In: *Quaderni di Sociologia*, XXIX (4), 595 - 639 (1981).
- [117] A. Martinez-Ruiz e T. Aluja-Banet. «Toward the definition of a structural equation model of patent value: PLS path modelling with formative constructs». In: *REVSTAT*, 7 (3), 265 - 290 (2009).
- [118] M. Mazziotta e A. Pareto. «Nuove misure del benessere: dal quadro teorico alla sintesi degli indicatori». In: *SISmagazine, rivista on-line della Società Italiana di Statistica*, <http://old.sis-statistica.org/magazine/spip.php?article194> (2011).
- [119] A. Monecke e F. Leisch. «semPLS: Structural Equation Modeling Using Partial Least Squares». In: *Journal of Statistical Software*, 48 (3) (2012).
- [120] G. Munda. *Multi-Criteria Decision Analysis and Sustainable Development*. Figueira, J. et al., 2005.
- [121] G. Munda. *Social multi-criteria evaluation*. Springer, Verlag, Heidelberg, New York, Economics Series, 2007.
- [122] G. Munda. «Social multi-criteria evaluation (SMCE): methodological foundations and operational consequences». In: *European Journal of Operational Research*, 158 (3), 662–677 (2004).
- [123] G. Munda e M. Nardo. «Constructing Consistent Composite Indicators: the Issue of Weights». In: Joint Research Centre, *Ispira* (2005).
- [124] G. Munda e M. Nardo. «Non-compensatory/Non-Linear composite indicators for ranking countries: a defensible setting». In: *Applied Economics* (2007).
- [125] J.H. Myers. *Measuring Customer Satisfaction hot button and other measurement issue*. Cengage Learning, 1999.
- [126] D. Nappo. «SEM with ordinal manifest variables An Alternating Least Squares approach». Tesi di dott. University of Naples Federico II, 2009.
- [127] M. Nardo et al. *Handbook on Constructing Composite Indicators: Methodology and User Guide*. OECD, 2005.

- [128] M. Nardo et al. «Tools for Composite Indicators Building». In: European Commission, *Ispira* (2005).
- [129] United Nations. *Human Development Report. United Kingdom: Oxford University Press*, <http://www.undp.org>, 1992, 1999, 2000, 2001.
- [130] R.G. Netemeyer, W.O. Bearden e S. Sharma. *Scaling Procedures: Issues and Applications*. Thousand Oaks, CA: *Sage Publications*, 2003.
- [131] H.H. Noll. «Social indicators and indicators systems: tools for social monitoring and reporting». In: *OECD, World Forum Statistics, knowledge and policy*. paper. Palermo, Italy, 2004.
- [132] R. Noonan e H. Wold. «Evaluating School Systems Using Partial Least Squares». In: *Evaluation in Education*, (7), 219 - 364 (1993).
- [133] OECD. *Quality Framework and Guidelines for OECD Statistical Activities*. OECD, www.oecd.org/statistics, 2003.
- [134] P. Paxton et al. «Monte carlo experiments: Design and implementation.» In: *A Multidisciplinary Journal*, 8 (2), 287 - 312 (2001).
- [135] S. Petter, D. Straub e A. Rai. «Specifying Formative Constructs in Information Systems Research». In: *MIS Quarterly*, 31 (1), 623 - 656 (2007).
- [136] K.J. Preacher e A.F. Hayes. «Asymptotic and resampling strategies for assessing and comparing indirect effects in simple and multiple mediator models». In: *Behavior Research Methods*, 40, 879 - 891 (2008).
- [137] K.J. Preacher e A.F. Hayes. «SPSS and SAS procedures for estimating indirect effects in simple mediation models». In: *Behavior Research Methods, Instruments, and Computers*, 36, 717 - 731 (2004).
- [138] Nilsson R. «Confidence Indicators and Composite Indicator». In: CIRET conference. paper. Paris, France, 2000.
- [139] R. Rajala e M. Westerlund. «Antecedents to consumers' acceptance of mobile advertisements – a hierarchical construct PLS structural equation model». In: *XLIIIth Hawaii International Conference on Systems Sciences (HICSS)*. paper. Hawaii, 2010.
- [140] F.F. Reichheld. «The One Number You Need to Grow». In: *Harvard Business Review*, 12 (2003).

- [141] W. Reinartz, R. Echambadi e W.W. Chin. «Generating non-normal data for simulation of structural equation models using Mattson's method». In: *Multivariate Behavioural Research*, 37 (2), 227 - 244 (2002).
- [142] W. Reinartz, M. Krafft e W.D. Hoyer. «The Customer Relationship Management Process: Its Measurement and Impact on Performance». In: *Journal of Marketing Research*, 41 (3), 293 - 305 (2004).
- [143] J. Rifkin. *The European Dream*. Tarcher-Penguin, New York, 2004.
- [144] E. Ringdon, C.M. Ringle e M. Sarstedt. *Structural modeling of heterogeneous data with partial least squares*. Malhotra N. (Ed) *Review of Marketing Research (Review of Marketing Research)*, 7, Emerald Group Publishing Limited, 255 - 296, 2010.
- [145] C. Ringle e R. Schlittgen. «A genetic algorithm segmentation approach for uncovering and separating groups of data in PLS path modeling». In: Oslo, Norway, 2007.
- [146] C.M. Ringle. «Segmentation for path models and unobserved heterogeneity: The finite mixture partial least squares approach». In: *Research Papers on Marketing and Retailing*, 35 (1), University of Hamburg (2006).
- [147] C.M. Ringle, S. Wende e A. Will. *Customer segmentation with FIMIX-PLS*. Aluja, T. et al., 2005.
- [148] C.M. Ringle, S. Wende e A. Will. *Finite Mixture Partial Least Squares Analysis: Methodology and Numeric Examples*. Esposito Vinzi, V., Chin, W.W., Henseler, J. e Wang, H. (Eds), *Handbook of Partial Least Squares (PLS): Concepts, Methods and Applications*, Springer, 195 - 218, 2010.
- [149] C.M. Ringle, S. Wende e A. Will. *SmartPLS 2.0 (beta)*. University of Hamburg, 2005.
- [150] R. Rosen. *Life Itself: A Comprehensive Inquiry into Nature, Origin, and Fabrication of Life*. Columbia University Press, 1991.
- [151] B.M. Russett. «Inequality and instability». In: *World politics*, 21, 422 - 454 (1964).
- [152] G. Russolillo. «Non-Metric Partial Least Squares». In: *Electronic Journal of Statistics*, 6, 1641 - 1669 (2012).
- [153] G. Russolillo. «Partial Least Squares Methods for Non-Metric Data». Tesi di dott. University of Naples Federico II, 2009.

- [154] T. Ryan e B. Joiner. «Normal probability plots and test for normality». In: Technical report, *Statistic Departement, The Pennsylvania State University, USA* (1976).
- [155] M. Saisana, A. Saltelli e S. Tarantola. «Uncertainty and Sensitivity Analysis Techniques as Tools for the Quality Assessment of Composite Indicators». In: *Journal of the Royal Statistical Society*, 168 (2), 1 - 17 (2005).
- [156] M. Saisana e S. Tarantola. «State-of-the-art. Report on Current Methodologies and Practices for Composite Indicator Development». In: *Institute for the Protection and Security of the Citizen Econometrics and Statistical Support to Antifraud Unit* (2002).
- [157] A. Saltelli. «Composite Indicators between analysis and advocacy». In: *Social indicators research*, 81, 65 - 77 (2007).
- [158] A. Saltelli e S. Tarantola. «On the relative importance of input factors in mathematical models: safety assessment for nuclear waste disposal». In: *Journal of American Statistical Association*, 97 (459), 702 - 709 (2002).
- [159] J. Salzman. «Methodological Choices Encountered in the Construction of Composite Indices of Economic and Social Well-Being». In: *Ottawa: Center for the Study of Living Standards* (2003).
- [160] G. Sanchez. *PLS Path Modeling with R*. Berkeley: Trowchez Editions, 2010.
- [161] G. Sanchez e T. Aluja. «A simulation study of PATHMOX (PLS path modeling segmentation tree) sensitivity». In: *Oslo, Norway*, 2007.
- [162] G. Sanchez e T. Aluja. *Pathmox: a PLS-PM segmentation algorithm*. Esposito Vinzi, V., Lauro, N.C., Braverma, A, Kiers, H. e Schmiek, M.G. (Eds) *Proceedings of KNEMO 2006*, Tilapia, Ancapri, 69, 2006.
- [163] G. Sanchez e T. Aluja. *Pathmox: Segmentation Trees in Partial Least Squares Path Modeling. R package version 0.1-1*. URL <http://CRAN.R-project.org/package=pathmox>, 2012.
- [164] G. Sanchez e L. Trinchera. *plsmp: Partial Least Squares Data Analysis Methods. R package version 0.2-2*. URL <http://CRAN.R-project.org/package=plsmp>, 2012.
- [165] F. Santini. *Le ricerche di mercato: come si fanno e come si utilizzano*. Luigi Buffetti, 1994.

- [166] M. Sarstedt. «A review of recent approaches for capturing heterogeneity in partial least squares path modelling». In: *Journal of Modelling in Management*, 3 (2), 140 - 161 (2008).
- [167] M. Sarstedt e C.M. Ringle. «Treating Unobserved Heterogeneity in PLS Path Modelling: A Comparison of FIMIX-PLS with Different Data Analysis Strategies». In: *Applied Statistics*, 37 (8), 1299 - 1318 (2010).
- [168] M. Sarstedt et al. «Uncovering and Treating Unobserved Heterogeneity with FIMIX-PLS: Which Model Selection Criterion Provides an Appropriate Number of Segments?» In: *Schmalenbach Business Review*, 63 (1), 34 - 62 (2011).
- [169] J.J. Schmid e J.M. Leiman. «The development of hierarchical factor solutions». In: *Psychometrika*, 22 (1), 83 - 90 (1957).
- [170] A. Sen. «Development as capabilities expansion». In: *Journal of Development Planning*, 19, 41 - 58 (1989).
- [171] A. Sen. «Development as Freedom». In: Oxford University Press, Oxford (1999).
- [172] A. Sen. «Poverty and Famines: An Essay on Entitlement and Deprivation». In: Oxford Clarendon Press (1981).
- [173] A. Sharpe. «Literature Review of Frameworks for Macro-Indicators». In: Ottawa: Center for the Study of Living Standards (2004).
- [174] G. Sánchez. «PATHMOX Approach: Segmentation Trees in Partial Least Squares Path Modeling». Tesi di dott. Universitat Politècnica de Catalunya, 2009.
- [175] M.E. Sobel. «Asymptotic confident intervals for indirect effects in structural equation models». In: *Sociological Methodology*, 13, 290 - 312 (1982).
- [176] D. Sörbom. «A general method for studying differences in factor means and factor structures between groups». In: *British Journal of Mathematical and Statistical Psychology*, 27, 229 - 239 (1974).
- [177] S. Squillacciotti. *Prediction oriented classification in PLS path modelling*. Aluja, T., Casanovas, J., Esposito Vinzi, V., Morineau, A. e Tenenhaus, M. (Eds), *Proceedings of PLS-05 International Symposium*, SPAD Test&go, Paris, 499 - 506, 2005.

- [178] M. Stone. «Cross-validatory choice and assessment of statistical predictions». In: *Journal of the Royal Statistical Society*, 36, 111 – 147 (1974).
- [179] A. Tanese, G. Negro e A. Gramigna. *La customer satisfaction nelle amministrazioni pubbliche*. Rubbettino Editore, 2003.
- [180] S. Tarantola, J. Jesinghaus e M. Puolamaa. *Global sensitivity analysis: a quality assurance tool in environmental policy modelling*. Saltelli, A., Chan, K. e Scott, M. (Eds) *Sensitivity Analysis*, John Wiley & Sons, New York, 385 - 397, 2000.
- [181] S. Tarantola et al. «Statistical techniques and participatory approaches for the composition of the European Internal Market Index 1992 - 2001». In: European Commission: JRC-Italy (2002).
- [182] S. Tarantola et al. «The Internal Market Index 2004». In: European Commission: JRC-Italy (2004).
- [183] A. Tenenhaus e M. Tenenhaus. «Regularized generalized canonical correlation analysis». In: *Psychometrika*, 76 (2), 257 - 284 (2011).
- [184] M. Tenenhaus. *La Régression PLS: théorie et pratique*. Paris: Technip, 1998.
- [185] M. Tenenhaus, S. Amato e V. Esposito Vinzi. «A global goodness-of-fit index for PLS structural equation modelling». In: *XLIIth SIS Scientific Meeting*. Contributed paper. Padua, Italy, 2004.
- [186] M. Tenenhaus, E. Mauger e C. Guinot. *Use of ULS-SEM and PLS-SEM to measure a group effect in a regression model relating two blocks of binary variables*. Esposito Vinzi, V. et al., 2010.
- [187] M. Tenenhaus et al. «PLS Path Modeling». In: *Computational Statistics and Data Analysis*, 48 (1), 159 - 205 (2005).
- [188] L. Trinchera. «Unobserved heterogeneity in structural equation models: A new approach to latent class detection in PLS path modeling». Tesi di dott. University of Naples Federico II, 2007.
- [189] L. Trinchera e V. Esposito Vinzi. «Capturing unobserved heterogeneity in PLS path modeling». In: Contributed paper. Ljubljana, Slovenia, 2006.
- [190] L. Trinchera, G. Russolillo e N.C. Lauro. «Using categorical variables in PLS Path Modeling to build system of composite indicators». In: *Statistica Applicata* (2008).

- [191] L. Trinchera, S. Squillacciotti e V. Esposito Vinzi. *PLS typological path modeling : a model-based approach to classification*. Esposito Vinzi, V. et al., 2006.
- [192] E.R. Trufte. *The Visual Display of Quantitative Information*. Graphic Press, Connecticut, USA, 2001.
- [193] E. Valdani. *Marketing strategico, un'impresa produttiva per sviluppare capacità market driving e valore*. RCS libri & Grandi Opere, 1995.
- [194] J. Van Rijckevorsel e J. de Leeuw. *An outline of PRINCALS*. Internal Report RB 002, Leiden, Department of Data Theory, University of Leiden, 1979.
- [195] S. Venaik. *A Model of Global Marketing in Multinational Firms: An Empirical Investigation*. Unpublished Doctoral Dissertation. The Australian Graduate School of Management, Sydney, 1999.
- [196] M. Wedel e W.A. Kamakura. *Market segmentation – conceptual and methodological foundations*. 2 edn. Boston: Kluwer, 2000.
- [197] M. Wedel e W.A. Kamakura. *Market segmentation: Conceptual and methodological foundations*. Kluwer Academic Publishers, Boston, 2000.
- [198] M. Wetzels, G. Odekerken-Schröder e C. van Oppen. «Using PLS path modeling for assessing hierarchical construct models: guidelines and empirical illustration». In: *MIS Quarterly*, 33 (1), 177 - 195 (2009).
- [199] B. Wilson. *Using PLS to Investigate Interaction Effects Between Higher Order Brand Constructs*. Esposito Vinzi, V., Chin, W.W., Henseler, J., Wang, H. (Eds), *Handbook of Partial Least Squares: Concepts, Methods and Applications in Marketing and Related Fields*, 2009.
- [200] B. Wilson e J. Henseler. «Modeling reflective higher-order constructs using three approaches with PLS path modeling: A Monte Carlo comparison». In: *Australian and New Zealand Marketing Academy Conference*, 791 - 800. paper. University of Otago, Dunedin, New Zealand, 2007.
- [201] R.S. Winer. *Marketing management*. Apogeo, 2002.
- [202] H. Wold. *Model construction and evaluation when theoretical knowledge is scarce*. Kmenta, J. e Rmsey, J.B. (Eds), *Evaluation of econometric models*, 47- 74, 1980.
- [203] H. Wold. «Modelling in complex situations with soft information». In: (1975).

- [204] H. Wold. *Partial Least Squares*. Kotz, S. e Johnson, N.L. (Eds), *Encyclopedia of Statistical Sciences*, 6, Wiley, New York, 581 - 591, 1985.
- [205] H. Wold. *PLS path models with latent variables: the nipals approach: the nipals approach*. Blalock, H.M. et al., 1975.
- [206] H. Wold. *Principles Estimation of principal components and related models by iterative least squares*. Krishnaiah, P.R. Editor, *Multivariate Analysis*, Academic Press, New York, 391 - 420, 1966.
- [207] H. Wold. *Soft modeling by latent variables: the nonlinear iterative partial least squares approach*. Gani, J. Editor, *Perspectives in probability and statistic*, Academic Press, London, 117 - 142, 1975.
- [208] H. Wold. *Soft modeling: the basic design and some extensions*. Jöreskog, K.G. e Wold, H. (Eds), *Systems under Indirect Observation: Causality, Structure, Prediction*, Part 2, North-Holland, Amsterdam, 1 - 54, 1982.
- [209] H. Wold, H. Martens e H Wold. *The multivariate calibration problem in chemistry solved by the PLS method*. Ruhe, A. & Kagstrom, B. (Eds) *Proceedings of the Conference on Matrix Pencils. Lectures Notes in Mathematics*, 1983.
- [210] F.W. Young, Y. Takane e J. de Leeuw. «The Principal Components of Mixed Measurement Level Multivariate Data: an Alternating Least Squares Method with Optimal Scaling Features». In: *Psychometrika*, 45, 279 - 281 (1978).
- [211] V.A. Zeithaml, A. Parasuraman e L.L. Berry. *Servire Qualità*. Mc.Graw-Hill, 1991.
- [212] P. Zhang, N. Li e H. Sun. «Affective Quality and Cognitive Absorption: Extending Technology Acceptance Research». In: *XXIXth Hawaii International Conference on Systems Sciences, IEEE Computer Society*. paper. Hawaii, 2006.
- [213] K.J. Zink. *Total quality management as a holistic management concept: The European model for business excellence*. Springer-Verlag, 1998.

Ringraziamenti

Grazie a tutti quelli che in questi anni mi sono stati vicini e, in vario modo, sono stati di stimolo e di aiuto nella stesura di questa tesi. Grazie alla Prof.ssa Grassia per avermi seguito in questo e nei precedenti percorsi formativi, per avermi scosso nei momenti in cui avrei voluto rinunciare e per avermi fatto crescere ancora un pò. Grazie a Rosanna che mi ha dato il suo supporto, mi ha infuso un pò di saggezza, mi ha dato ottimi consigli e mi ha mostrato la strada da intraprendere in questo lungo percorso, da buona ricercatrice e ottima amica. Grazie a Gabriella, a lei il merito di avermi riportato sulla strada della ricerca e di avermi spronato ad iniziare questo percorso. Grazie a Giuliana, Elena e Valeria che mi hanno aiutato a portare a termine la stesura di questa tesi con i gruppi di studio nel Weekend. Grazie a tutti i ragazzi del XXX e del XXIX ciclo per avermi reso parte di una grande famiglia, purtroppo sono stato poco presente, ma questi anni sono stati più leggeri grazie a voi. Grazie a gli amici di sempre, ai miei coinquilini e ai nuovi amici di questi anni che mi hanno fatto da spalla e con cui ho condiviso giornate e sarate spensierate. Grazie a Stefano che mi ha dato la possibilità di intraprendere questo percorso di crescita personale e culturale, supportatandomi in questa scelta, da ottimo capo e da grande persona. Grazie a Gennaro, Chiara, Pia e a tutti i colleghi, che rendono le giornate a lavoro più gradevoli. A Pingu, che meno male che c'è. Grazie ai miei genitori, che sempre mi sorreggono e mi aiutano. Un doveroso ricordo va infine alla Prof. Balbi, a lei che è sempre stata presente nel mio percorso di formazione, e che è sempre stata esempio di passione e professionalità.